MACRO-LEVEL COLLISION AND CRIME ANALYSIS: CASE STUDY FOR THE CITY OF REGINA

A Thesis Submitted to the College of
Graduate and Postdoctoral Studies
In Partial Fulfillment of the Requirements
For the Degree of Master of Science
In the Department of Civil, Geological and Environmental Engineering
University of Saskatchewan
Saskatoon

By
EMMANUEL A. TAKYI

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make this thesis freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis. I also reserve the right to use in any future works (such as articles or books) all or part of this thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Civil and Geological Engineering
57 Campus Drive
University of Saskatchewan
Saskatoon, Saskatchewan S7N 5A9 Canada

ABSTRACT

Traffic collisions and crimes are issues of concern in most neighbourhoods and cities and they are certainly a concern for the City of Regina. The traditional approach to either preventing or reducing the severity of collisions and crimes has been a reactive one: identifying locations as problematic based on historical data before taking action. An advanced and recently introduced approach for dealing with the issues of collision and crime is the Data-Driven Approaches to Crime and Traffic Safety (DDACTS). DDACTS is a proactive, place-based approach that identifies problematic locations that require interventions. Results from a macro-level analysis are used for planning purposes.

Traffic Analysis Zones for the City of Regina were considered in this research. Traffic Analysis Zones are a spatial aggregation of census blocks and are, in part, a function of population, used by city planners, for planning new neighbourhoods and resource allocation, as well as by transportation officials for tabulating traffic-related data. Traffic Analysis Zones level collision and crime prediction models have been developed to estimate safety and security effects of neighbourhood level land use, socio-economic factors, road network characteristics, and demographic variables on collisions and crimes. Furthermore, the Empirical Bayes technique are adopted to estimate expected frequencies of collisions and crimes. The expected frequencies are used in determining hotspots that require enforcement and countermeasures.

The Negative Binomial modeling technique was adopted in this study to predict numbers of collisions and crimes. Models were calibrated and validated using multiple goodness-of-fit tests. Results from the goodness-of-fit tests were used as basis to determine the best model for predicting each type of collision and crime. Maps were then created to display both spatial patterns and spatio-temporal trends of collisions and crimes. Traffic Analysis Zones with significant frequencies of collisions and crimes, both separately and in unison, were then identified.

Some of the conclusions drawn from the collision prediction models include: both intersection density and intersection road density had positive associations with collisions; and when comparing 3-leg and 4-leg intersections, 3-leg intersections had fewer safety concerns. Also, low density residential areas have collision reduction effects. Results from collision prediction models developed in this study can help transportation engineering officials, and city planners in traffic safety decision. At the planning stage of new neighbourhoods, the safety effects of

individual predictors or sets of predictors can be determined by creating multiple scenarios that involve interested sets of variables.

The developed crime models provided information about how land use type, socio-demographics, and residential land use type influence different crime types. Some conclusion drawn include the following: commercial areas and retail spaces were target areas for high numbers of violent crimes; high population density neighbourhoods attracted high numbers of crimes; higher numbers of residents within the age groups of 18 to 24 and 25 to 44 were positively associated with both violent and non-violent crimes; residents within the age groups of 44 to 65 as well as 65 years and over had a crime reduction effect, regardless of the crime occurrence type. Also, low density residential areas attracted many non-violent crimes; industry and office areas also attracted many non-violent crimes; and multiple or mixed land use areas also attracted a high volume of auto-involving theft crimes.

The results of this research is intended to improve the lives of the residents of the City of Regina by providing tools that can be used to reduce traffic collisions and crimes.

ACKNOWLEGDMENTS

I thank God Almighty for the strength, guidance and protection throughout this journey. His grace abounds always!

I thank my parents for their unflinching support in prayers, in kind and in encouraging words, your counsel guided me throughout this research. Mr. Billy Elvis, Mr. Ali Iddris and Ms. Emelia Ashie; you are very much appreciated.

My sincere appreciation goes to Dr. Peter Y. Park for his guidance and supervision during this research. Your professionalism in helping me complete this research is very much appreciated. I have learnt a lot of lessons that can be applied in both academia and life in general.

I also will like to express much gratitude to my research committee members Dr. Kerry Mazurek, Dr. Amin Elshorbagy, and Dr. Avi Akkerman for their valuable contribution in completing this research.

I also thank Seun Oluwajana, my research colleague, with whom I worked tirelessly, for his contribution. I thank Miriam Okadigbo, Yaw Darko Offei-Darko, Charlene Nortey, George Ato Eguakun, and Kayla for their contributions towards my thesis. The support of George Nii Ashie, Godfred Nii Ashie, Diana Croffie, Dennis Marfo and Glenn Ross during this research is also appreciated.

Finally, I will like to thank the funding agencies for this research; Saskatchewan Government Insurance (SGI), Regina Police Service, Saskatchewan Centre of Excellence for Transportation and Infrastructure (SCETI), and the City of Regina for their contribution towards this research.

DEDICATION

This research is dedicated to my grandmother, Felicia Adwoa Takyiampong and Edna Abena Dabiwaa Okrah!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AADT- Average Annual Daily Traffic
AIC- Akaike Information Criterion
BAC- Blood Alcohol Content
BIC- Bayesian Information Criterion
COR- City of Regina
CPM- Collision Prediction Model
CSI- Crime Severity Index
FI- Fatal-Injury
GIS- Geographic Information Systems
GOF- Goodness of fitness
HVE- High Visibility Enforcement
MAD- Mean Absolute Deviation
MPB- Mean Prediction Bias
MSPE- Mean Squared Prediction Error
MSE- Mean Squared Error
NB- Negative Binomial
PDO- Property Dame Only
R2FT- Freeman Turkey R-Squared
RMSE- Root Mean Squared Error
RPS- Regina Police Service
TAZ- Traffic Analysis Zones
TLKM- Total Lane Kilometers
VKMT- Vehicle Travel Kilometers

CHAPTER 1 . INTRODUCTION

1.1 Problem Statement

In 2013, Saskatchewan had the highest collision casualty rate of 12.6 per 100,000 population compared to an average 5.5 per 100,000 for Canada (Canadian Motor Vehicle Traffic Collision Statistics, 2013). In 2012 and 2013, the City of Regina, which has a population of 215,004, had sixteen traffic fatalities, and that was the highest compared to other cities in Saskatchewan (2013 Saskatchewan Traffic Accident Facts, 2014). From 2009 to 2013, the City of Regina experienced approximately 20% (6,054) of the total number of collisions (29,411) as fatal-injury collisions (Saskatchewan Government Insurance [SGI], 2014). Fatal-injury collisions are a combination of fatal and injury collisions. Fatal collisions result in at least one person sustaining bodily injury and resulting in death within 30 days of the date of the collisions (Capital Region Intersection Safety Partnership, 2012). Injury collisions result in at least one person sustaining an injury but not leading to death (Capital Region Intersection Safety Partnership, 2012). Figure 1.1 shows the steady trend of total collisions in Regina over the five-year period (2009-2013).

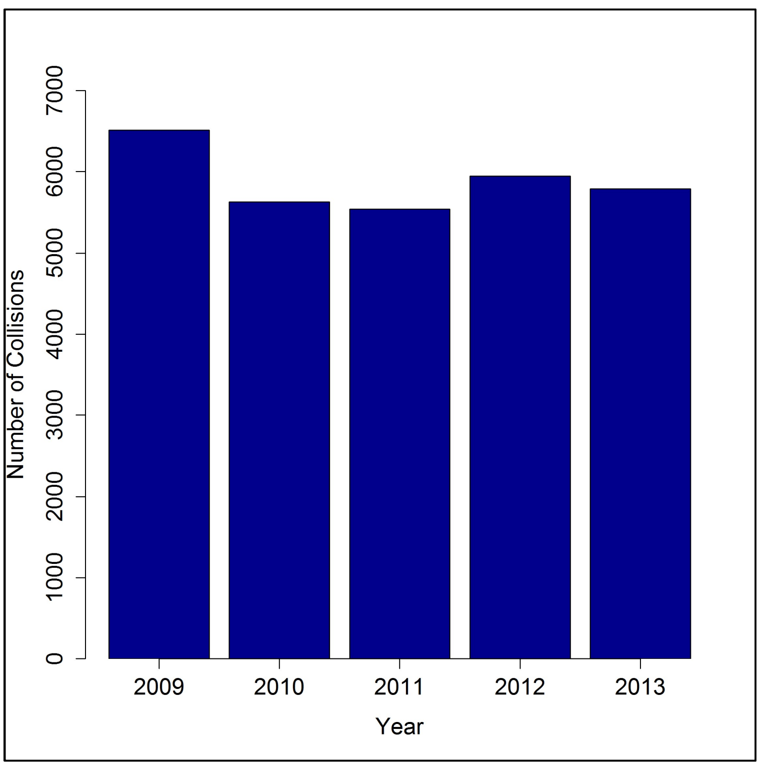


Figure 1.1: Yearly Collision Frequency for the City of Regina (2009-2013)

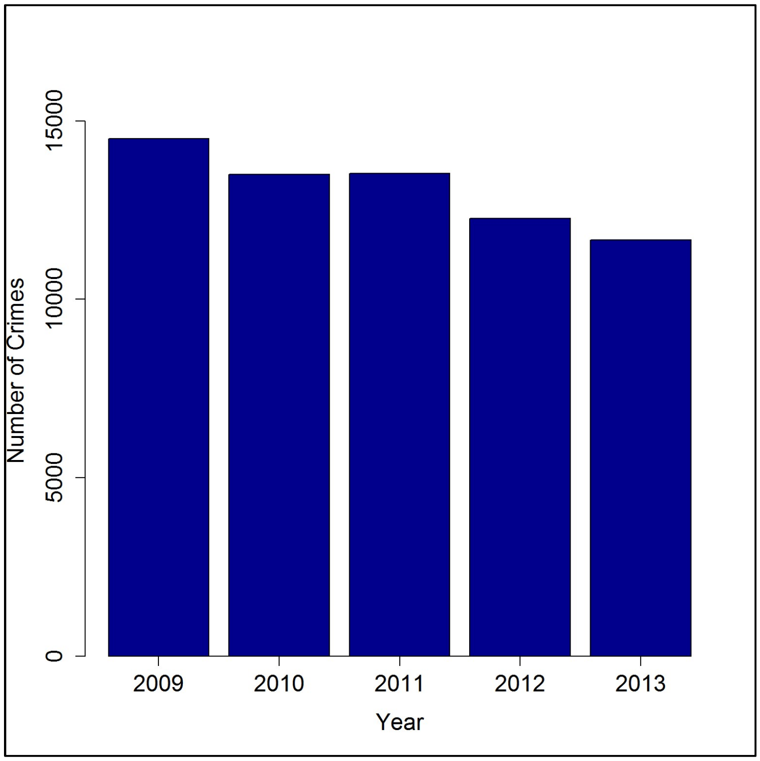


Figure 1.2: Yearly Crime Frequency for the City of Regina (2009-2013)

In 2013, the City of Regina recorded the highest Crime Severity Index of 109.3 per 100,000 population compared to an overall 69 per 100,000 population for Canada; the City of Regina's highest crime rate of 8,069 per 100,000 population compares to 5,191 per 100,000 for Canada (Boyce *et al.*, 2014). The Crime Severity Index is an index to compare reported crimes for different places. All crimes are assigned some form of weight based on their seriousness. For instance, murder will have the highest weight compared with theft under $5,000 and break and enter crimes. This index is based on the sentence handed to individuals who commit such crimes. Total crimes reported are then weighted based on their seriousness and the summation represents the Crime Severity Index. Therefore, murder and other serious crimes have a significant impact on Crime Severity Index calculation and changes in CSI. (Statistics Canada, 2009). Figure 1.2 shows the trend for crimes for the City of Regina over the study period (2009-2013), and illustrates a decreasing trend of the total number of crimes (Regina Police Service, 2014).

The trends for high occurrences for both collisions and crimes have been similar over the study period of this research (2009-2013). Similar trends were also identified for impaired driving collisions and violent crimes. Figures 1.3 and 1.4 represent clockplots that are designed to show the peak hours of impaired driving collisions and violent crimes, respectively. The rings represent days of the week and the different dot sizes represent frequencies of collisions and crimes. The straight lines converging at the centre represent hours in a day from 1:00 am to 12:00 midnight. Bigger red dots represent higher numbers of incidents, green dots represent lower numbers of incidents, and the orange and yellow dots represent collision frequencies between the red and green. The plots show distribution of impaired driving collisions and crimes during the days of the week (Monday-Sunday) and hours within which they occurred. From Figures 3 and 4, it can be seen that the majority of incidents occur between the hours of 23:00 and 4:00, especially on weekends.

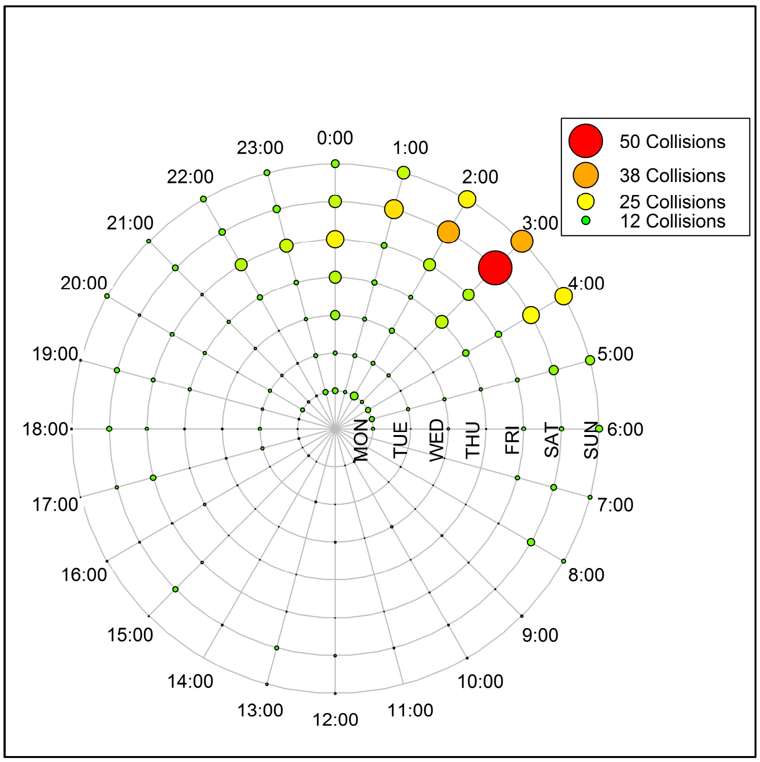


Figure 1.3: Impaired Driving Collisions by Hour and day of the Week (2209-2013)

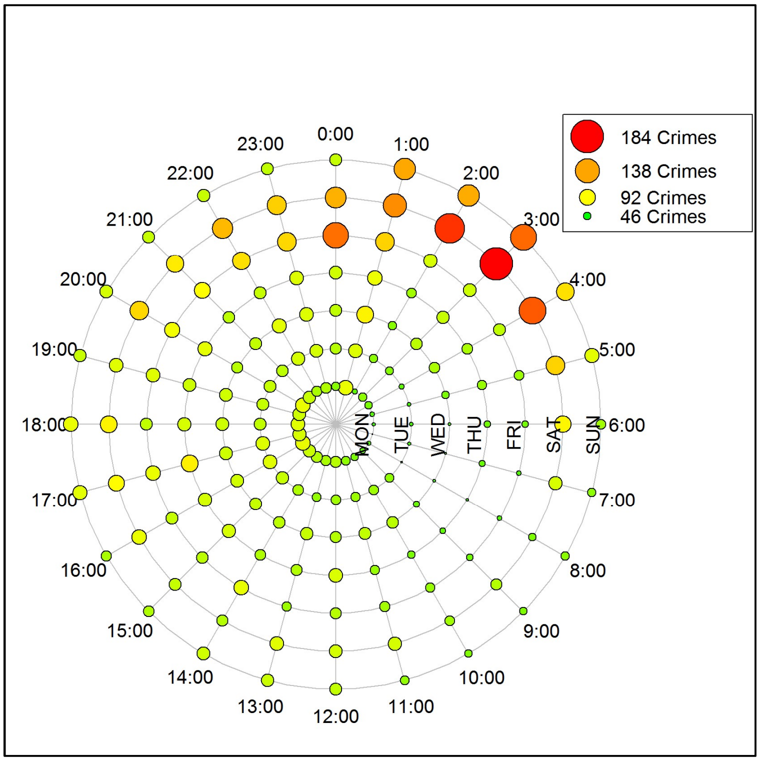


Figure 1.4: Violent Crimes by Hour and Day of the Week (2009-2013)

To reduce the frequency and severity of the occurrence of both crimes and collisions, some countermeasures can be adopted by various agencies. These can be grouped into four main areas:

- Education;
- Engineering;
- Emergency Medical Services; and
- Enforcement.

The first area, education, is the use of outreach programs to inform, guide, and warn the public about rules and safe driving practices to prevent or reduce the occurrence of traffic collisions and crimes. Education involves the use of mass media publicity, such as television advertisements or billboards, to create awareness about both collisions and crimes with the intention of reducing the occurrence of both incidences. Insight-based educational programs for young drivers are an important component of educational countermeasures (Bates *et al*, 2006). Some educational interventions adopted in Canada include Graduated License Education, Report Impaired Driving, bicycle handling, and proper visibility by bike riders (Bates *et al*, 2006; SGI, 2010 and City of Thunder Bay, 2010). An observational study of one educational intervention, "concluded that there was no evidence of an interaction between the effects of speed camera ticketing and speed-related publicity awareness on the frequency of casualty crashes" (Saskatchewan Government Insurance, 2005; Saskatchewan Government Insurance, 2010; Mayhew & Simpson, 2002). Instead, "the effect of speed-related publicity during 1996-2000 was due to advertising with emotive styles" (Cameron *et al*., 2003). That implies educational measures that implore feelings such as emotions of people are effective in reducing traffic collisions. Studies involving three different data sources, including individual-level data from the census on incarceration, Uniform Crime Reporting of state-level arrests data, and self-report data on crime and incarceration acquired through the National Longitudinal Survey of Youth, showed similar conclusions when addressing criminal violations: schooling significantly reduces criminal activity (Lochner *et al*., 2003).

Engineering provides countermeasures, which can be either vehicular or roadway, and are geared towards reducing traffic collisions. Various technological features have been designed by automotive industries to help drivers operate vehicles in a safe manner to prevent and reduce the severity of collisions. These advanced technologies have been designed for specific functions: front collision prevention, adaptive cruise control, lane departure warning and lane keeping warning system, blind spot detection, park assist and back over prevention, adaptive curve headlights, fatigue warning systems, electronic stability control, and antilock brakes. Various research and studies have been conducted to identify the reduction effects of some advanced technologies designed for vehicles and found that vehicles with advanced technology, such as electronic stability control system, forward collision warning, adaptive headlights, advance emergency braking system, as well as pedestrian and city automatic braking systems, had 16% fewer claims under property damage liability coverage compared to vehicles without the technology (Highway Loss Data Institute, 2012).

There are various engineering roadway countermeasures that prevent and reduce the severity of collisions. Countermeasures include rumple strips, guard rails, and median barriers to prevent overtaking and prevent head-on collisions. Road humps, raised pedestrian crossings, reflectors, and chevron signs, and advanced advisory speed limits for sharp curves are also included. Roadway interventions have shown a significant reduction in frequency and severity of traffic collisions. Chen *et al*. (2012) examined the relative safety effects of pedestrian countermeasures at urban intersections using a case study in New York. The pedestrian countermeasures included increasing the cycle length, Barnes Dance, traffic signal split phase timing, traffic signal installation, and a high visibility crosswalk. Chen *et al*. (2012) performed a before-and-after treatment study, and the results showed there was a 50% reduction in the average number of pedestrian-involved collisions per intersection by increasing the cycle length; a 51% reduction in the average number of pedestrian-involved collisions per intersections by adopting Barnes Dance; a 39% reduction in average number of pedestrian-involved collisions per intersection by implementing a signal split phase timing; a 12% reduction in the number of pedestrian-involved collisions by installing signals; and a 40% reduction in the average number of pedestrian-involved collisions by using a high visibility crosswalk.

Emergency Medical Services, the third area of countermeasures, are mostly post-collision interventions to reduce fatalities and/or seriousness of injuries as a result of collisions. These services involve medical personnel that have been trained with on-field skills to respond quickly to and treat injured persons involved in traffic collisions. Emergency services are time sensitive and the most important factor is a quicker response to incidents. It is mostly a pre-hospital care aimed at avoiding preventable death and disability, limiting the severity and suffering caused by the injury and ensuring optimal function of collision survivors and reintegration into the community (World Health Organization, 2004). Some of the emergency response services involve determination of hotspots or high collision areas and pre-deploying emergency vehicles to such hotspots to reduce travel time to incidents (Estochen *et al*., 1998). The safety effects of pre-deployment of emergency vehicles to collision hotspots were estimated by Estochen *et al* (1998) using GIS in before-and after-scenarios. Pre-deployment is strategically parking emergency service vans at locations other than hospitals and deploying those vehicles to respond to traffic collisions sites. Using a six-year (1990-1995) collision data from the Iowa's Accident Location and Analysis System for Des Moines, Iowa, Estochen *et al* (1998) employed GIS techniques to determine response times to collisions for five, seven, and ten minutes response times. These response times were then compared to response times for the current locations of emergency service vehicles. The analysis showed that pre-deployment of emergency vehicles decrease response times for many incidents: within 5 minutes, 63.2% of total collisions would be responded to compared to 56.4% at the current location; within 7 minutes, 83.8% of total collision would be attended to compared to 81.9% at the current location; and within 10 minutes, 94.5% of total collisions would be attended to compared to 94.7% at the current location (Estochen *et al*., 1998). In the case of 10 minutes response times, traffic collisions that were responded to reduced by 0.2%, which in comparison with 5 and 7 minutes response times is significant. Emergency services are therefore, vital in reducing the severity of collisions.

Finally, enforcement involves members of society, mainly law enforcement agencies, enforcing laws by identifying, preventing, and punishing persons who violate society's governing laws. Using technology, such as speed-cameras, breathalyzers, drug test kits, and automatic license plate recognition, enforcement is an effective and important measure to reduce the frequency and severity of both crimes and collisions (Royal Canadian Mounted Police, 2015; Shimizu & Desrochers, 2015). Enforcement can be used to target specific lifestyles that contribute to traffic fatalities or injuries, and programs can be geared towards the safety of pedestrians, disabled personnel, bicycle riders, motorbike riders, and drivers.

An example of an enforcement action includes requiring a helmet for all motorized two-wheel users to prevent head injuries, which is the main cause of death among riders. One program adopted across many jurisdictions in North America, including Saskatchewan, is the Selective Traffic Enforcement Program. Selective Traffic Enforcement Program involved the collaboration of Saskatchewan Government Insurance, Saskatchewan Justice, and Saskatchewan police communities to conduct enforcements, targeted at certain traffic violations as well as collisions (Government of Saskatchewan, 2016). Selective Traffic Enforcement Program showed positive effects, such as seat belt usage increasing from 59% to 90% from 1986 to 2000, and a reduction of alcohol impaired fatal collisions from 35-50% to 28.5% over 20 years of implementation of Selective Traffic Enforcement Program (Government of Saskatchewan, 2016).

Banning the use of hand-held devices, and implementing traffic speed reduction enforcement and construction cone zone enforcement are some of the countermeasures to reduce the frequency and severity of collisions as well as traffic-related crimes such as manslaughter through dangerous driving, criminal negligence by street racing and criminal negligence by street racing causing injury or death (Saskatchewan Government Insurance, 2013). Over the past four decades in Canada, technological advancements in the engineering design of vehicles and other countermeasures have contributed to significant reduction in road fatalities by a factor of three, despite the doubled population (Transport Canada, 2011).

The main focus of this research is identifying locations that can benefit from enforcement as a tool to reduce the frequency and severity of collisions and crimes. Traditionally, police enforcement tactics are usually reactive: response to 911 calls for service by the public; sporadic patrols by police officers with the intention of running into a crime or traffic offense scene; community policing and police officers waiting until a location is identified as problematic before taking law enforcement measures. In contrast, proactive enforcement systems include sophisticated intelligence-led policing, predictive policing, and selective and smart enforcement. Data-Driven Approaches to Crime and Traffic Safety (DDACTS) is one of the innovative and advanced tools for enforcement (National Highway Traffic Safety Administration, 2014).

1.2 Data-Driven Approaches to Crime and Traffic Safety (DDACTS)

Recently, a new enforcement tactic, known as Data-Driven Approaches to Crime and Traffic Safety (DDACTS), is being introduced to many police services in North America as a more effective way to control traffic violations, traffic collisions, and crimes. Regarded as one of the most advanced law enforcement models, DDACTS is a law enforcement operational model that uses an integration of location-based traffic and crime data to maximize the effectiveness and efficiency of available resources. It is a proactive and effective tool for reducing crime and preventing future crime, while simultaneously minimizing traffic collisions. DDACTS is a place-based policing, which targets locations of interest such as commercial areas, residential areas *etc*. as opposed to a person-based policing which targets people of a certain ethnicity or culture, and studies have shown that place-based policing is a more efficient law enforcement model (Weisburd, 2008).

Researchers in criminology and criminal justice have fueled research investigating the relationship between crimes, traffic collisions/violations, and place-based policing. A study was conducted to examine the relationship between traffic fatalities and crime. The study indicated that traffic fatalities result from incivility and aggression, demonstrating a neglect for social conventions, and other serious violations like homicide (Giacopassi and Forde, 2000). In 1997, police in Albuquerque, New Mexico introduced a Safe Streets program, involving saturation patrols, follow-up patrols, highway speed enforcement, and sobriety checkpoints. This program was developed after identifying 27 of 33 high-collisions locations in four general geographic areas, which were also identified as high-crime areas. This program showed significant results: a 9% decrease in property damage collisions, an 18% decline in injury collisions, a 20% decline in impaired driving collisions, a 34% decline in fatal collisions, a 29% decline in homicides, a 17%

decline in kidnapping, and a 10% decline in assaults (Stuster, 2001). This reduction in both collision and crime is significant compared to previous year's data.

A long-term study of juvenile crimes during a 14-year period (1989-2002) indicated that 50% of all juvenile crimes occurred in less than one percent of Seattle's street segments. All juvenile crime incidents occurred at less than five percent of street segments (Weisburd *et al.*, 2011). This finding supports the theory that certain crime types occur in close proximity to roads.

Furthermore, increased police traffic enforcement patrolling showed a reduction in traffic collisions and in certain crimes, such as property losses and personal injuries (Schnelle *et al.*, 1977). Traffic enforcement between 1994 and 1996 in Peoria, Illinois showed significant results: a 24% increment in traffic citations, a 28% officer initiated activity, a 16% custodial arrests, an 11% driving under influence (driver whiles impaired) arrests, a 21% reduction in traffic collisions, a 6% reduction in citizen generated calls, a 12% decrement in part one crime index, a 10% decrement in violent crimes, and a 12% decrement in property crimes (National Highway Traffic Safety Administration, 1997). This finding also supports the theory of the effect of traffic enforcement on crimes.

A study involving 119 vehicular homicides, crimes that involve the death of an individual other than the driver due to criminal negligence or murderous operation of a vehicle, showed that victims and offenders were similar to violent crimes offenders: "the tendency towards aggressive behaviour, characteristic of a subculture of violence, influences the way an individual drive as well as his face-to-face interactions" (Michalowski, 1975). This study also supports the theory that personnel committing collisions are also likely to commit certain types of crimes.

Studies in geographic criminology have shown that specific places attract individuals who commit certain crimes and traffic violations. These findings are supported by the deviant place theory, which suggests that victims themselves do not encourage crime, but the high-crime neighbourhoods they reside in put them at risk of coming into contact with criminal offenders (Stark, 1987). Juvenile crime was focussed in public and commercial areas where youth gather—schools, youth centres, shops, malls, and restaurants—rather than residential areas (Weisburd *et al.*, 2011). Therefore, places of overlapping occurrence of both crimes and collisions become very likely and more predictable.

Because of DDACTS's focus on the importance of place, DDACTS is increasingly becoming a preferred enforcement approach. First using evidence-based problem-solving, DDACTS utilizes location-specific data collected over a period of time, and such interventions are geared towards problematic areas based on historical data. In such areas, the implementation of high visibility enforcement is effective, as such enforcement advocates for the use of police officers in official uniforms and the use of marked police vehicles. The presence of enforcement officers in problem areas deters offenders from committing traffic offences, as well as crimes. Since DDACTS is a location-based enforcement tactic and targets a geographic location, it eliminates legal and ethical concerns that result from targeting an ethnic group or persons of a particular descent. Finally, the utilization of a predictive—rather than reactive—method targets problem locations. Therefore, incidents are prevented from occurring rather than responded to after they have occurred. As illustrated, evidence-based problem solving, high visibility enforcement, less ethical and legal concerns, and predictive method are significant influences for the increasing enforcement of DDACTS.

DDACTS as an operational level law enforcement model relies on seven guiding principles listed below (National Highway Traffic Safety Administration, 2014):

1. Partnership and stakeholders' participation;
2. Data collection;
3. Data analysis
4. Strategic operations
5. Information sharing and outreach
6. Monitoring, evaluation, and adjustments; and
7. Outcomes.

This study focusses on the second and third DDACTS guiding principles. Data collection involves collecting accurate and timely collision, crime, and enforcement related data. Data analysis is the creation of actionable analysis results including maps that overlay collision and crime data that allows agencies to identify hotspots (National Highway Traffic Safety Administration, 2014).

1.3 Research Goal and Objectives

The goal of this research is to develop a data-driven analysis to identify problem areas or hotspots where significant numbers of collisions and crimes occur, and overlap for the City of Regina, using the concept of DDACTS. The main objectives for this research are the following:

1. To develop Traffic Analysis Zone level collision and crime prediction models for the City of Regina, using an advanced statistical technique;
2. To develop a geographic information system (GIS) based collision and crime mapping system, to display existing and predicted numbers of collisions and crimes; and
3. To identify hotspots where significant numbers of crimes and collisions occur and overlap to show where and when Regina Police Service should focus its law enforcement program.

1.4 Benefits of Research

Over the study period of this study, the City of Regina had the second highest rate of criminal coded crimes in Canada, and one-third of crimes are related to motor vehicles (Boyce *et al*., 2014). This research is expected to contribute in decision making aimed at reducing the loss of life and property that are caused by collision and crime incidents. The hotspot maps created with the GIS tool will help improve the allocation of Regina Police Service and Saskatchewan Government Insurance budget for law enforcement. By reducing known and unknown flaws in the current approach, the new system of enforcement can be expected to deliver substantial economic benefits to the society.

The predicted collisions and crime data for the City of Regina can be applied in various decisions, including safety improvements to sections of the road network, improvements to Traffic Analysis Zones (TAZ), future planning and expansion of the city, distribution of infrastructure, and allocation of resources. Also, the collision prediction models produced from this research can be applied to road safety engineering projects in the City of Regina. Similarly, the crime prediction model can provide some valuable input for the Regina Police Service's future enforcement tactics. This research can be applied in other jurisdictions across North America and other parts of the world to achieve a common goal of reducing societal harm by reducing traffic collisions and crimes.

1.5 Scope

The area of study is the City of Regina. Using collision and crime historical data (2009-2013), projected socio-economic data, and road network basemaps, this research will focus on the development of temporal-spatial hotspot maps. The five-year historical traffic collision data represents the minimum data requirements for predicting collision outlined in the highway Safety Manual. Employing the concept of DDACTS, this research will identify hotspots for areas of overlapping high incidents of crimes and collisions. Furthermore, by using peak hour data, hotspots for peak hours for both crimes and collisions will be identified. Prediction models are to be developed for both collisions and crimes and hotspot maps will be developed for the predicted incidents and hotspots identified.

Issues related to modifiable areal unit and aggregate data assignment are beyond the scope of this study and will not be discussed. Solutions to these issues will be identified through literature review and such solutions will be adopted in this study.

1.6 Layout of Thesis

Chapter Two of this thesis contains a literature review of collision and crime prediction, reduction programs, and prediction models. Next, Chapter Three discusses the research data and descriptive statistics used for the data. Chapter Four describes the methodology used in both collision and crime prediction models. Chapter Five presents the modelling results and traffic analysis zones that were identified as hotspots. Finally, Chapter Six presents the conclusions and recommendations of this thesis.

CHAPTER 2 . LITERATURE REVIEW

This chapter provides a literature review related to road safety, transportation planning, collision prediction models, and crime prediction models. This literature review aims to describe the research context and theoretical foundations on which this research is based.

2.1 Collision Prediction Models

Collision prediction models are regression models used in road safety to estimate the number of collisions to be expected in a transportation network. Collision Prediction Models can be site-specific—including intersection, road segment, highway, ramps, and terminals—or can be on a regional and zonal level, using units such as Traffic Analysis Zones. Because it is a spatial aggregation of census blocks, a Traffic Analysis Zone, in part, is a function of population (Peters and MacDonald, 2004). A Traffic Analysis Zone is an area demarcated by state and/or local transportation officials for tabulating traffic-related data, especially trip generation and attraction statistics, and is defined as part of the Census Transportation Planning Package (United States Department of Transportation Federal Highway Administration, 2007). Collision Prediction Models are developed by establishing the relationship between a dependent variable, collision frequency, and several independent or explanatory variables. By incorporating large numbers of variables and their relationships, Collision Prediction Models explain differences in collision frequency. Using this statistical modeling, early transportation researchers sought to determine the systematic relationships that exist between traffic collisions, traffic volume, roadway geometry, and infrastructure.

Conventional linear regression was adopted for predicting traffic collisions; however, over time, it was realised that conventional linear regression was inappropriate for modeling traffic collisions. First, traffic collisions are random events and can fluctuate considerably. Collisions are discrete and non-negative and, therefore, the normally distributed error structure of conventional linear models makes them inappropriate to predict traffic collisions. However, Poisson and Negative Binomial regression models have proven to be appropriate for modeling traffic collisions (Lovegrove and Sayed, 2006).

2.1.1 Micro-Level Collision Prediction Models

As an operational level Collision Prediction Model, micro-level collision prediction models represent the most basic level at which collisions can be predicted. Micro-level Collision Prediction Models, therefore, predict the number of expected collision for a specific road segment or a specific intersection within a road network over a specified period of time (Highway Safety Manual, 2010). Collision prediction models are also referred to as Safety Performance Functions. Micro-level Collision Prediction Models, are used for both safety improvements and countermeasures aimed at improving the safety and performance of a transportation network (highway Safety Manual, 2010). Such improvements and countermeasures include but are not limited to the installation of a traffic signal, addition of exclusive right or left turn lanes, traffic calming measures, and installation of median or road barriers. Thus, micro-level network screening is a reactive or retrofit approach, requiring years of collision data to identify hotspots that require safety improvements.

2.1.2 Macro-Level Collision Prediction Models

The Macro-level Collision Prediction Model approach complements the traditional reactive approach, which has been in practice for decades. Macro-level Collision Prediction Models are on a much larger unit of analysis and are used at the planning stage. The results from a macro-level Collision Prediction Model are used as tools in decision-making for future planning, provision of infrastructure, and safety evaluation of existing road network (Hadayeghi *et al*., 2003; Hakim *et al*., 1991; and Lovegrove & Sayed, 2005) . Consequently, this level of analysis has a much larger area of analysis that combines road segments and intersections. The area of analysis could be a community, neighbourhood, or an even larger area, depending on the purpose for which the analysis is intended. This collision model is also known as a zonal collision prediction model. This proactive approach aims at preventing collisions from occurring by prioritizing road safety at the design and planning stage of road networks and communities.

Macro-level collision prediction models have been conducted by researchers on various aggregated levels, such as neigbourhood, Traffic Analysis Zone, county, and even provine-wide levels. Such works are discussed in this section. For instance, using monthly-aggregated data between January 1974 and December 1986, a macro-level Collision Prediction Model was developed for 18 Norwegian counties by Fridstrøm and Ingebrigtsen (1991). Collision Prediction Models were developed for fatal and injury collisions. The authors considered a wide range of potential explanatory variables: exposure (gasoline sales, 1000 bus kilometers driven in scheduled transport), weather (snow or rain), daylight (minutes of daylight during rush hours, minutes of daylight at night [midnight sun]), road network (length of road, real fixed capital kilometer national road, annual maintenance), collision reporting practices, vehicle inspection, law enforcement, usage of seat belts, proportion of inexperienced drivers, and alcohol sales. Poisson regression models were employed in their analysis. Next, they identified the issues of heteroscedasticity and collinearity, suggesting that Poisson or negative binomial techniques inherently takes into account the heteroscedasticity but not the autocorrelation of collision frequencies. Heteroscedasticity is the assumption that the errors or residuals associated with a model are unequal across the range of the dependent variable (Laffont *et al*, 2004). Collinearity also known as multicollinearity, on the other hand, refers to the existence of high correlation among the independent variables in a model. The existence of heteroscedasticity and collinearity among variables affect the predictive performance of models.

Fridstrøm *et al.* (1995) modeled monthly aggregated data on a county level from four countries: Denmark (14 counties), Finland (11 counties), Norway (19 counties) and Sweden (24 counties). Poisson regression was employed using two variable groups: weather and exposure. During their analysis, five goodness-of-fit tests were employed to select the best model: log-likelihood ratio, the overdispersion parameter (Elvik index), multiple correlation coefficient (R^2), weighted R^2, and Freeman-Turkey transformation residuals (R^2_{FT}). Brännäs and Johansson (1992) studied the same dataset for Sweden used by Fridstrøm *et al*. (1995) to check for autocorrelation in the data. The review showed that a significantly high autocorrelation existed among variables used in their analysis. Fridstrøm *et al*. (1995) explained that estimates obtained by Brännäs and Johansson (1992) employed negative binomial and that the differences in their results were insignificant and negligible.

In 1949, Smeed used data from 1930 to 1936, including an aggregated population, the registered number of motor vehicles, and traffic fatalities for macro-level analysis. Using country as the unit of analysis, Smeed (1949) sought to determine a statistical relationship between fatalities, population, and the number of registered motor vehicles. Smeed (1949) used aggregated data for 20 countries. Great Britain, Northern Ireland, Eire, United States, Australia, Canada, South

Africa, New Zealand, Denmark, Finland, Norway, Sweden, Belgium, France, Netherlands, Italy, Germany, Portugal, Spain and Switzerland. After analysis, the research derived a positive linear relationship between collision fatalities, population, and the number of registered motor vehicles. Smeed (1949), however, stressed that the positive relationship does not consistently hold and further deduced that the higher the registered number of motor vehicles per population ratio, the greater the death rate per unit population. This study which used countries as the unit of analysis may suffer from Modifiable Areal Unit Problem because of the very larger level of aggregation. Modifiable Areal Unit Problem are errors that occur when one groups data into a unit of analysis (Heywood *et al*, 1998). Grouping and generalizing data across large unit of analysis such as country does not reflect the true nature of the data for different parts of the country.

Using collision data from 1980-1994 for 41 of the 56 countries in the Asia Pacific region, Kumara and Chin (2004) studied fatal collision using the Negative Binomial technique. The results of this research showed that traffic fatalities increased with population, number of registered vehicles, per capita Gross National Product, and road length. The analysis also showed that pacific regions were associated with a higher numbers of fatalities. The number of fatalities, however, reduced with time. This study also likely suffered from the issue of Modifiable Areal Unit Problem die to the large area of analysis used.

Also, in reviewing macro-level Collision Prediction Models, Hakim *et al.* (1991) reviewed 14 publications by various authors. The review compared variables with coefficients that were statistically significant at the probability level of 80% and above. In their research, Hakim *et al*. (1991) considered fatal collisions and injury collisions. After reviewing these publications, Hakim *et al.* (1991) emphasized issues to be considered when developing a macro-level collision prediction model. First, the use of collision frequency as the dependent variable is preferred to collision rate (collision per capita). Also, cross-sectional data (e.g. county or Traffic Analysis Zone) should be used instead of time-series data (e.g. monthly). The other issue identified was collinearity among variables, as collinearity makes it difficult to assess the effect of individual variables on collision frequency. Collinearity is the existence of high correlation between two independent variables in a model. The implication of collinearity is that the one independent variable can predict the other independent variable in the model. Variables such as economic, socio-demographic, and driving-related variables exhibit high collinearity. The existence of these collinearities possibly affected the predictive capability of the models developed.

Lovegrove and Sayed (2006) developed macro-level Collision Prediction Models to evaluate neigbourhood traffic safety. Using aggregated neigbourhood-level data, such as traffic volume, demographics, network variables, and transportation demand variables, Lovegrove and Sayed (2006) studied 577 neighbourhoods in the Greater Vancouver Regional District. The Generalized Linear Model, assuming a Negative Binomial error structure, was adopted. The Collision Prediction Models revealed that increased numbers of collisions were associated with increases in the following variables: total transit and vehicle kilometers travelled, total lane kilometers, average congestion level, total commuters from each zone, shortcut capacity on local roads through zone, shortcut capacity with average congestion level used to adjust, number of drivers commuting from zones, commuter density, signal density [number of signals per hectare], intersection density [number of intersections per hectare], percentage of arterial-local intersections per total lane kilometers, percentage of arterial lane kilometers per total lane kilometers, workers per resident [number of workers/population], population density, unemployed workers in total labour force, and home density. Lovegrove and Sayed (2006) explained the increase in collisions associated with increase in Vehicles Kilometers Travelled, Total Lane Kilometers, average congestion level, and number of drivers commuting from zones confirm their expectations and suggested that increasing number of signalized intersections may not be necessarily safer. On the other hand, the number of collisions reduced with increase in the number of average family size, core area [max area without major roads], core area as a percentage of total zonal area, percentage of three-way intersections per number of intersections, and percentage of local lane kilometers per total lane kilometers variables.

Lovegrove and Littman (2007) developed macro-level Collision Prediction Models using 479 neighbourhoods in the Greater Vancouver Regional District to evaluate the road safety effect of mobility management strategies, known as traffic demand management. Traffic demand managements are policies and programs developed to reduce traffic and parking congestion, and pollution emissions from vehicles. Traditionally, road safety is not the main objective of traffic demand managements. The safety effects of the traffic demand management on fourteen variables were studied and the variables that had significant impact were used in their macro-level Collision Prediction Model. The conclusion from their research was that in traffic demand management; transportation and land use factors have the potential to increase road safety in addition to the conventional objectives, such as environmental, social, and economic benefits.

Pulugurtha *et al.* (2013) developed collision estimation models using 2005 data and 1057 Traffic Analysis Zones for Charlotte, North Carolina. Collision data, land use data, street centre line network, and a Traffic Analysis Zone layer with embedded planning data in GIS were used for this analysis. Pulugurtha *et al.* (2013) observed that population, number of household units and employment, trip production and trip attraction, and centre-lane miles by speed limit were highly correlated to land use data characteristics; thus, such variables were not used in the development of models. Collision data obtained included 24 fatal collisions, 3522 injury collisions, and 7180 Property Damage Only collisions for the 2005 study period. Even though there were 1057 Traffic Analysis Zones in the study area, 24 were excluded in the modelling process because those areas were open land area. Of the data, 65% was randomly selected for model calibration, and the remaining 35% was used for model validation. Correlations between the independent variables were investigated; as a result; socio-economic, network data, urban residential commercial, rural district, and mixed-use district were excluded from their modeling. Negative binomial regression modeling technique was adopted and number of collisions were predicted for: total number of collisions, injury collisions, and Property Damage Only collisions. The models were validated by the chi-squared statistic. Models were rejected if the chi-squared statistic was lower than the critical 95% confidence level. All three models to predict total, injury and Property Damage Only collisions passed the chi-squared statistics test.

Hadayeghi *et al.* (2003) developed macro-level Collision Prediction Models to evaluate safety effects of urban transportation systems, using 463 Traffic Analysis Zones for the City of Toronto. Negative Binomial was employed in their modeling. Models were developed for total collisions, fatal collisions, and property damage only collisions. Variables used include socio-economic, demographic, traffic demand, and road network data. Resulting from their analysis, the variables that were significant in collision prediction were number of households, major road kilometers, vehicle kilometers traveled, intersection density, posted speed, and volume-capacity ratio. Hadayeghi *et al.* (2003) further developed a collision prediction model for morning peak hours (6:00 am – 9:00 am) and found similar results. Table 2.1 shows a summary of the effect of the model variables for models developed. A positive (+) effect means an increase in that variable increases the number of estimated collisions, and a negative (-) effect means an increase in that variable results in a reduction in the number of estimated collisions. The effects were the same for total collisions and morning peak hours, irrespective of the severity.

Table 2.1: Variables and estimated effect (Hadayeghi *et al.*, 2003)

Total Collisions	
Variable	Estimated Effect (+) or (-)
Natural log of vehicle kilometers traveled	(+)
Major Road Kilometer	(+)
Number of Households	(+)
Posted Speed	(-)
Volume/Capacity	(-)
Intersection Density	(+)

Macro-level collision studies were also performed by Noland (2003), Noland and Oh (2004), and Noland and Quddus (2004). In these three studies, analysis was done at a much higher level of aggregation. In one analysis, Noland (2003) used country as the analysis level; in the other, Noland and Oh (2004) considered county-level of Illinois; and, in the last, Noland and Quddus (2004) identified a ward-level of the Great Britain. These analyses are likely to suffer a phenomenon known as modifiable areal unit problem. Modifiable areal unit problem, errors are created when data are grouped or aggregated into one unit or multiple units for analysis, results in either distortion or exaggeration of the actual data. Modifiable areal unit problem is beyond the scope of this research and as such will not be discussed in detail.

Noland (2003), studied the contribution of medical and technological advances to the reduction of fatalities in developing countries. The author observed 10 years (1978-1989) of road and collision data from the International Road and Traffic Accident Database, as well as health data from the Organization for Economic Cooperation and Development. Using the Negative Binomial modeling technique, Noland (2003) estimated the number of fatalities per country. Due to a lack of data, some countries were excluded from the analysis, making interpretation of results a challenge. The estimated effects of the independent variables also varied significantly for different models. Despite these conditions, the author concluded that advances in medical treatment are likely contributing factors to reduction in traffic fatalities.

Noland and Quddus (2004) developed macro-level models to predict the numbers of fatalities, serious injuries, and slight injuries, as well as motorized and non-motorized fatalities. The authors used 1999 collision data for 8,414 wards in the Great Britain. They found that fatalities were associated more with rural wards than urban wards. Furthermore, wards with lower population density and higher unemployment density had fewer fatalities. The presence of intersections had no effect on fatalities. However, length of roadways slightly affected serious injuries. In this study, employment and population were normalized by area of wards, and the normalized variables were used as proxies for traffic exposure. Another finding from this research indicated models for slight injuries tend to overestimate the predicted injuries. However, because the study used only one-year data, the analysis may have suffered regression-to-mean bias. Regression-to-the-mean bias is a statistical tendency that a site with an extreme frequency in one year is likely to have a less extreme collision frequency the following year (American Association of State Highway and Transportation Officials, 2010). Hadayeghi *et al.* (2006) sought to explore the temporal transferability of zonal level collision prediction models. Results from earlier research, which used zonal level data from 1996 to 2001 to develop Collision Prediction Models for the City of Toronto, was updated to measure the performance of the model for later years. Calibrated data from 1996 was used to predict collisions for 2001. This study used traffic collisions, sociodemographic factors, road network characteristics, and traffic flow for 463 (1996) and 481 (2001) Traffic Analysis Zones for the City of Toronto. The study's results of the effects of variables are summarized in the Table 2.2.

Table 2.2: Comparison between esimated variable effects for 1996 and 2001 Collision Prediction Models (Hadayeghi *et al*., 2006)

Variables	Sign of estimates	
	1996	2001
Total major road length (km)	(+)	(+)
Number of households (x 10^{-3})	(+)	(+)
Speed	(-)	(-)
Volume/Capacity ratio (v/c)	(-)	(-)
Intersection density	(+)	(+)

However, the results from the updated 1996 calibration model to predict 2001 collisions were inconsistent with the developed Collision Prediction Model for 2001 data. This finding suggests that temporal transferability of zonal level Collision Prediction Models, regardless of whether the spatial unit is the same, is inaccurate and unreliable. Expanding on this research, Hadayeghi (2009) developed zonal level safety planning models and examined their temporal transferability using data from 461 Traffic Analysis Zones in 1996 and 481 Traffic Analysis Zones in 2001 for the City of Toronto. As an extension of the work discussed earlier by Hadayeghi *et al*. (2006), the inference was consistent with the earlier research: temporal transferability of zonal level Collision Prediction Models is not feasible.

Naderan and Shahi (2010) developed what they called Aggregated Crash Prediction Models at Traffic Analysis Zone level for Mashhad, Iran. The city consisted of 380 Traffic Analysis Zones, and the data used for this analysis were traffic collisions and trip generations. The authors sought to predict collisions using trip production and attraction for the various Traffic Analysis Zones. They concluded that trip production and trip attractions based on work, shopping, or school showed a positive effect on the frequency of collisions. This effect can be explained by the fact that most of these trips were made with personal vehicles and during rush hours. Moreover, trip production and attraction based on educational and recreational purposes showed a negative effect on collision frequency. The authors argued that these findings were influenced by the fact that such trips were made outside of rush hours. Furthermore, the short proximity of schools to residences led people to prefer other alternatives, such as walking or using public transit.

Similar research was conducted by Abdel-Aty *et al.* (2011), using 1349 Traffic Analysis Zones of four counties in Florida, US to predict severe, peak hour, pedestrian, bicycle-related, and total collisions using trip attractions and productions as well as roadway characteristics. Traffic collisions for 2005 and 2006 and thirteen different types of trip attractions and productions were considered in this research. The outcome of their analysis showed that roadways with posted speed limits of 35, 45, 55, and 65 miles per hour (mph) showed a positive association with severe collisions. However, roadway lengths with posted speed limits of 25 mph were negatively associated with collisions. Also, roadway lengths with posted speed limits of 45 and 65 showed the highest association with total collisions. Moreover, home based work productions, home based shop attractions, and home based other productions showed positive associations with peak hour collisions. Intersection density (intersection per Traffic Analysis Zone) showed a positive association with every type of collision investigated in their research.

Ihssian (2014) investigated the influence of boundary data assignment on the development of multimodal macro-level collision prediction models, using 422 Traffic Analysis Zones for the City of Ottawa. Traffic collisions are usually recorded in databases on road segments or intersections; therefore, in aggregate level analysis, assignment of collisions to various zones based on their geographical location is required. Similar to the way some intersections may be a boundary for multiple Traffic Analysis Zones, some road segments are also boundaries for multiple Traffic Analysis Zones. Ihssian (2014) adopted ten different boundary data assignment methods, and the various methods showed significant influence on the predictive capability of models. These 10 methods included assignment based on population, population-employment ratio, equal-proportion, total lane kilometer, vehicle-kilometer traveled, and multiple-count methods. Collision Prediction Models were created for different levels of collision frequency and types: total, injury, Property Damage Only, bike-involved, and pedestrian-involved collisions. In conclusion, Ihssian (2014) found that an even distribution of boundary collision data between adjacent Traffic Analysis Zones was the best method since it significantly improves the predictive capability of the models created. One of the key observations with this research was that the number of traffic signals in a Traffic Analysis Zone was positively correlated with all the different types of collisions.

Wang and Huang (2016) researched road network safety evaluation by employing the Bayesian hierarchical joint model with micro-level and macro-level analysis. The authors argued that road safety is a microscopic problem; therefore, contributing factors are micro-level in nature, and macro-level Collision Prediction Models do not address the real problem. However, this claim is weak and does not address the actual application of macro-level Collision Prediction Models: they are intended to be used as a planning level tool and not for making immediate decisions. Nonetheless, the focus of Wang and Huang's (2016) research was to relate collision at the road network level (road segments and intersections) to macro-level (Traffic Analysis Zone) variables and micro-level road network characteristics by using Meso Collision Prediction Models. Their approach was to avoid the issue of boundary collision assignment. Using 544 road segments and intersections that were clustered into the 208 county level Traffic Analysis Zone, data included collision data, 16 micro-level (segment and intersections) variables, and four macro-level (Traffic Analysis Zone) variables. The outcome of this research showed that comparisons between micro-level Collision Prediction Model, macro-level Collision Prediction Model, and the joint model revealed a much improved predictive performance by the joint model. Such findings suggest that a joint model is an innovative approach to predicting collision due to the Traffic Analysis Zone level data, which are associated with planning or development.

Huang *et al*. (2016) compared micro-level Collision Prediction Models and macro-level Collision Prediction Models. A hot zone identification method was adopted to clearly define the advantages of each approach. Collision data from 2005-2007 were obtained from the Crash Analysis Reporting in Florida. The collisions were then aggregated into 155 Traffic Analysis Zones in Hillsborough County, Florida. After their analysis, they concluded that because micro-level models showed a better fitting and superior predictive capability, the models can be employed to suggest countermeasures for problem areas. Huang *et al*. (2016) concluded that macro-level analysis required much less detailed data. Furthermore, the outcomes from such models provide non-traffic engineering issues and a powerful tool in developing long term transportation plans.

2.1.3 Application of Macro-Level Collision Prediction Models

Analysis from macro-level Collision Prediction Models can be used in transportation planning and quantifying the safety benefits of the following:

- Regional transportation plan (Lovegrove., Lim., and Sayed, 2010)
- Evaluating the road safety effects of mobility management strategies (Lovegrove and Litman, 2007)
- Changes in speed limits (Grabowski and Morrisey, 2007; Lave, 1985; Rock, 1995; Shafi and Gentilello, 2007)
- Regional enforcement (Yannis, Papadimitriou, and Antoniou, 2007)
- A reduction in the maximum Blood Alcohol Content (Kaplan and Prato, 2007)
- Differences in rural and urban fatality and hospitalization rates (Kmet and Macarthur, 2006)
- Evaluating neigbourhood traffic safety (Lovegrove and Sayed, 2006)
- Economic growth (Kopits and Cropper, 2005)
- Changes in socio-demographics and infrastructure (Noland and Oh, 2004)
- Seatbelt use and related legislation (Majumdar, Noland, and Ochieng, 2004)
- Medical treatment and improvements in technology (Noland, 2003; Noland and Quddus, 2004)
- National safety campaigns (Van Schalwyk, 2000)
- Medical facilities in rural areas (Street, Winter, Buckley, Nicholson, and Twomey, 1999)
- Network safety management (Burrow and Taylo, 1995); and
- Unemployment (Leigh and Waldon, 1993)

2.2 Network Screening

As outlined earlier, data analysis step of DDACTS is one of the focuses of this research. Data analysis involves using statistical and visual tools to identify areas with concern. Analyzing and identifying areas of concern can be termed as network screening. Network screening is the process of reviewing a transportation network to identify and rank sites with the potential of benefiting from a safety improvement. This step also identifies locations with a particular collision type or severity. Various measures are used for network screening, including collision frequency, collision rate, and expected collision frequency, using a collision prediction model. A Collision Prediction Model equation estimates or predicts the expected average collision frequency per year at a location. Collision Prediction Model is essentially, a function of traffic exposure and, in some cases, road network characteristics, such as number of lanes, traffic control, or type of median (Highway Safety Manual, 2009). Extensions of roadway characteristics or infrastructures are the number of intersections, speed limits, roadway length, road class, and various derivatives of these characteristics.

Identifying sites that require safety improvements is an important step in the road safety management process. One such technique is the hotspot approach, which has been adopted in various agencies in safety improvements. A hotspot is a location within a road network that has a very high collision potential and, thus, has the potential for safety improvements. Site-specific collision potential is determined using measures such as collision frequency, collision rate, and collision severity, or a combination of any of the above-mentioned measures. There are, however, several issues with these measures, and they can lead to incorrectly identifying sites as hotspots.

Collision frequency is defined as the number of collisions recorded (observed) at a location during a specific time period. Collision frequency also known as collision counts is the simplest of techniques and requires few years of data, usually three years. However, the observed collision frequency at a location does not reflect the true safety due to the highly random nature of traffic collisions. Typically, time periods between one and three years are used to minimize the effects of random fluctuations and sensitivity to changes over time (Hadayeghi, 2009).

2.3 Crime Prediction Models

Crime predictions over the past decades have been either qualitative or quantitative. Quantitative crime prediction is briefly discussed, and a literature review on quantitative crime prediction are examined in this section. Quantitative crime prediction involves statistical tools and advanced mapping techniques to correlate past crime trends with various predictors, such as sociodemographic, socio-economic, roadway infrastructure, and neighbourhood characteristics to predict crimes. This technique is particularly important in forecasting the future scope of criminal activity (Schneider, 2002).

Osgood (2010) researched the advantages of modeling aggregated robbery crime data for juveniles (aged 10-17 years) using negative binomial regression. To demonstrate how useful Negative Binomial analysis would be for aggregated crime data, the results from five separate models, which used the same data, were compared. Osgood (2010) collected five-year data (1989-1993) for 264 non-metropolitan counties across Florida, Georgia, South Carolina, and Nebraska. Explanatory variables for these counties were provided by the United States Department of Commerce. Table 9 summarizes the data used for this study.

Table 2.3: Crime Variables and their description (Osgood, 2010)

Variable	Description
Residential instability	Proportion of household occupied by persons who had moved from another dwelling in the previous five years
Ethnic Heterogeneity	Proportion of house occupied by white versus non-white persons
Family disruption	Proportion of female-headed households with children
Poverty	Proportion of persons living below the poverty line
Unemployment rate	Proportion of unemployed persons as a ratio of workforce
Proximity to metropolitan counties	A dummy variable with 1 being adjacent to metropolitan statistical area, and 0 not being nonadjacent.
Population at risk	Youth aged between 10 and 17 years

The five modeling techniques were the following: Ordinary Least-Squares regression analysis for crime rate; Ordinary Least Squares for crime rate plus 1 (log(crime rate + 1)); Ordinary Least Squares for crime rate plus 0.2 (log(crime rate + 0.2)); basic Poisson regression; and Negative Binomial regression analysis. Osgood (2010) used four statistical goodness-of-fit tests to compare the performance of all five models: mean squared error; Pearson's R^2; 2 times the log likelihood for Poisson; and Negative Binomial and the Spearman R. Mean squared error is an indication of how close predicted numbers are close to the observed numbers. Mean squared error measures the average of the squares of errors associated with a predictor (Washington *et al*., 2005). Pearson's R^2 measures how the difference or variation in one can be used to explain the variation in a second variable (Glenn, 2012). The maximum likelihood estimation is used to determine the probability distribution that makes an observed data most likely (Myung, 2003). In other words, the maximum likelihood is a method that seeks to identify values of parameter vectors that maximise the likelihood function for an observed data. Spearman R, similar to Pearson's R, assesses how the relationship between two variables can be explained by a monotonic function Glenn, 2015). Concluding that the Negative Binomial regression was the best analysis technique for aggregated crime data, Osgood (2010) recommended adopting Negative Binomial technique for such analysis.

Pratt (2001) assessed the relative effects on macro-level predictors of crime. The author completed an extensive review of existing macro-level crime analysis research work conducted between 1960 and 1999. In this research, Pratt (2001) used different levels of aggregation and they included neighbourhood, Census tract, city, county, standard metropolitan statistical area, state, country, and multi-level. Using variables that are suggested by some seven theories of crime for macro-level analysis, Pratt's (2010) study revealed that racial heterogeneity (calculated as the proportion of white or non-white), poverty, and family disruption are very strong predictors of crimes.

To identify factors most important in crime prediction, Schneider (2002) reviewed and examined crime prediction works conducted in the first two decades of the 21st century. During the study, demographic characteristics were cited as one of the strongest predictors of crimes. Males aged between 15 and 25 were the predominant age group that had a high association with crimes; consequently, neighbourhoods with high populations of males aged between 15 and 25 had higher numbers of crimes. Another observation was that the strength of the economy had a high association with crime: periods with economic recessions showed a rapid increase in property crimes, whereas lower crimes were recorded during economically favourable times. Furthermore, Schneider (2002) observed that another significant variable that would influence future crimes was technology.

Cohen and Gorr (2005) developed crime forecasting techniques and mapping for the US police by using 103 square grid cells 4000 by 4000 feet (approximately 10 blocks) of Pittsburgh, Pennsylvania as the analysis unit. Nine-year (1990-1998) crime data was used. Economic and socio-demographic information obtained for each grid cell was also acquired from the US Census.

Various researchers have found the correlation between some land use types and violent crimes. Such results confirm natural intuition. For instance, areas with many social centres, such as pubs and drinking bars, often attract some violence. Stucky and Ottensmann (2009) further explored the correlation between land use and violent crimes and whether they are affected by "socioeconomic disadvantage." Crime data from 2000-2004, 2002 land use data, and 2000 Census data, were used in this study. The units of analysis were 2,142 1000 x 1000-feet grid cells, and the Negative Binomial technique was adopted in modeling the data. Aggregated violent crimes constituted murder and non-negligent manslaughter, robbery, aggravated assault, and rape. Crime counts per cell were determined by assigning crimes to cells based on geocodes that came with the acquired crime data. Unknown location cases were excluded from analysis. Socioeconomic, as well as demographic data, were geocoded for each cell.

Based on previous research that showed positive associations between crimes and poverty, unemployment, and family disruption, Stucky and Ottensmann (2009) developed a disadvantage index. The index included the following descriptions: proportion of poor people, proportion of unemployed persons, median household income, and proportion of female-headed households. Another variable, which was derived was the stability index, included proportion of owner-occupied households, proportion of foreign born persons, and proportion of persons that have not moved in the past five years. The disadvantage index would be used as an indicator for the economic disadvantage. Certain variables were used as dummy variables (1, if present; otherwise, 0) and these included high-density residences, cemeteries, schools, and hospitals. Major road lengths were also calculated and aggregated for each cell. The analysis by Stucky and Ottensmann (2009) showed that spatial lag, cell population, disadvantage index, high-density residences, and major roads total length were positively associated with violent crimes. However, the proportion of residential land use and proportion of industrial land use were negatively associated with violent crimes. Moreover, when some variables in the models were conditioned by the disadvantage index, it showed some differences in their significance, as well as counts for violence crimes.

Machin *et al.* (2010) researched the effect of education on crime in Britain. Their analysis showed that higher levels of education were negatively associated with criminal activity. Various compulsory school leaving age laws were used in their analysis. In their study, a strong negative association existed between property crimes and higher education; therefore, they concluded that higher education has social benefits and should be considered a main tool in policy change in crime reduction programs. The authors argued that education has a crime reduction effect due to the fact that education increases how much a person earns. Thus, any time educated, employed people spent outside of their regular working hours is costlier and, therefore, they are less likely to commit crimes. Furthermore, Machin *et al.* (2010) argued that time spent at schools limits the available time for teenagers to engage in criminal activities.

Shingleton (2012) researched violent crime trend prediction using regression analysis for the city of Salinas, California. This study used crime data received from the Salinas Police Department, as well as environmental factors, unemployment data, prison statistics, and police budgetary assignment. A correlation matrix between individual variables and violent crimes was created. Variables with high correlations with violent crimes were selected as candidate variables for the following: model and included population, Salinas Police Department budget, sworn police with a one-year shift, proportion of overpopulation in California Department of Corrections and Rehabilitation, proportion of unemployment, parole population, and personnel per household with a two-year shift. A further correlation matrix was developed among those variables; population and parole population and were removed from further analysis. Three regression techniques were employed: Ordinary Least Squares, Poisson, and Negative Binomial. Although all three models had predictions that were higher than actual counts, Ordinary Least Squares had predictions closer to the actual counts. Because the model provides a better fit, the author concluded Ordinary Least Squares is a better modeling technique for crime data.

Heim (2014) visualized and modeled crime data indexed by road segments. Macro-level data, including age, gender, population, housing prices, and characteristics, as well as police calls for service, were reassigned to road segments. The road segments were the units of analysis for this research work. Different statistical modeling techniques were employed, including the zero-inflated Negative Binomial and Poisson-Gamma Conditional Autoregressive techniques. Analysis revealed that calls for service, social disorders (e.g. complaints, noise violations), and prices of houses were the most significant variables in predicting total crimes.

In an attempt to explore the importance of small area unit analysis compared to a much larger aggregated unit of analysis, Wheeler (2015) compared the effects of using different sizes of areas as units of analysis. The goal was to identify aggregation bias caused by data aggregation.

Wheeler (2015) argued that crime prediction at street level is more practical for police enforcement but added that the unit of analysis chosen should be based on the researcher's intended results. Therefore, he chose street segments as the units of analysis and modeled with the Negative Binomial technique, concluding that the number of liquor stores on a street segment doubles the annual frequency of part 1 crime (homicide, sex offense, robbery, assault with a deadly weapon, larceny, burglary, stolen auto, theft from auto, and arson).

2.4 Modeling Techniques

Over the past few decades, the most common techniques for modeling traffic collisions have been either the conventional linear regression model or generalized linear regression assuming a Poisson or Negative Binomial error structure.

2.4.1 Linear Regression Models

Various researchers have attempted to predict collisions using a linear regression model by establishing the relationship between collisions and traffic volumes (Ceder and Livneh, 1982; Ceder, 1982). In a linear regression model, the dependent variable is predicted from a number of independent variables using a linear equation (Robert Nau, 2014). Equation 2.1 is the functional form of linear regression model (Nau, 2014):

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots.. + \beta_k x_{ik} \qquad 2.1$$

Where Y_i is the predicted number of collisions (response variable); x_{i1}, x_{i2}, x_{i3}, ……. x_{ik} are the predictors (explanatory variables); and β_0, β_1, β_2, β_3, …… β_k are the parameters (coefficient) estimates from the linear regression model.

To apply the Linear Regression Model as a modelling technique, the following assumptions must be valid:

- Errors associated with variables must be normally distributed;
- A linear relationship should exist between independent and dependent variables;
- There is no auto-correlation among variables;
- There is no or little multicollinearity; and
- The assumption of homoscedasticity, which implies that the errors or residuals associated with a model are equal across the range of the dependent variable (Laffont *et al*., 2004).

If these assumptions are not met, the results from such a model will be either over-estimated or under-estimated. Traffic collisions as events, are rare and random in nature and as such the number of incidents can fluctuate significantly over a short period of time. Due to these properties and the fact that traffic collisions are non-negative and discrete, conventional linear regression model cannot be used. This is because, conventional linear regression models assume a normally distributed error structure. The limitations associated with linear regression modeling collision data is overcome by generalized linear regression assuming Poisson or negative binomial error structure

2.4.2 Poisson Regression

Poisson regression assumes that the dependent variables in a regression analysis are counts that follow the Poisson distribution and that the observations are independent of the expectation (Jovanis and Chang, 1987). In this research the expectations are the number of collisions or crimes. This basic assumption of the Poisson distribution implies that the number of collisions or crimes occurring within an observed time interval (hours, days, years *etc*.), is independent of the expected number of collisions or crimes. The expected number of collisions within a time interval is a function of multiple variables such as the traffic volume, road length, roadway speed limit *etc*. As such, the expected numbers of collisions vary from time to time and this is known as a nonstationary Poisson process. The occurrence of traffic collisions and crimes can therefore be described as a nonstationary Poisson process. The model form for a nonstationary Poisson process is determined by Equation 2.2 (Jovanis and Chang, 1987):

$$\lambda_i = f(\beta, x_i) \qquad 2.2$$

where λ_i is the expected number of incidents (collisions or crimes) for the ith time interval, β is the vector of parameters to be estimated, and X_i is the vector of the independent variables for the ith time interval.

In a Poisson regression, the probability of a site (*e.g.* Traffic Analysis Zone, intersection or a road segment) i, having y_i collisions per year is determined by (Kim *et al*., 2010):

$$P(y_i) = \frac{\exp(-\lambda_i)\,\lambda_i^{y_i}}{y_i!} \qquad 2.3$$

$P(y_i)$ is the probability that site i, has y_i collisions per year, and λ_i is the Poisson parameter of site i. λ_i is equal to the expected number of collisions, $E(y_i)$, at site i. λ_i is a function of the predictor variables. The simplest form of the relationship for the Poisson parameter and a predictor variable is given by:

$$\lambda_i = exp(\beta x_i) \qquad 2.4$$

Where x_i is the predictor variable and β is an estimate from the Poisson model. The coefficient vector β can be estimated by the maximum likelihood function (Chang, 2005):

$$\prod_i \frac{exp[-exp(\beta x_i)][\beta x_i]^{y_i}}{y_i!} \qquad 2.5$$

Moreover, in Poisson regression, a satisfying property is that the variance and mean are the same. If this condition is not satisfied, the data is said to be either under-dispersed (where mean is greater than variance) or over-dispersed (when variance is greater than mean), and regression results would be biased.

2.4.3 Negative Binomial

The Negative binomial also known as Poisson-gamma is an extension of Poisson regression model introduced primarily to overcome the issue of overdispersion with Poisson regression. Overdispersion can be caused primarily by omitting variables that influence the Poisson rate (Kim *et al*., 2010). To overcome the issue of overdispersion, in Negative Binomial modeling, an error term, ε_i, is introduced in a Poisson model. This ε_i violates the assumption that the mean of collision frequencies is equal to the variance. Rewriting Equation 2.4 with the error term, ε_i, becomes:

$$\lambda_i = exp(\beta x_i + \varepsilon_i) \qquad 2.6$$

The error term, ε_i has a mean of one and a variance α, and $exp(\varepsilon_i)$ is a Gamma function. This results in a conditional probability function:

$$P(y_i|\varepsilon) = \frac{exp[-\lambda_i exp(\varepsilon_i)][\lambda_i \exp(\varepsilon_i)]^{y_i}}{y_i!} \qquad 2.7$$

Integrating ε out of Equation 2.7 results in an unconditional distribution of y_i. The resulting functional form of Negative Binomial will be (Chang, 2005):

$$P(y_i) = \frac{\Gamma(\theta + y_i)}{[\Gamma(\theta) \cdot y_i!]} \cdot {u_i}^{\theta}(1 - u_i)^{y_i} \qquad 2.8$$

where $u_i = \frac{\theta}{\theta + \lambda_i}$, $\theta = \frac{1}{\alpha}$, α is the variance of the gamma-distributed error term, and $\Gamma(\theta)$ is a value of the Gamma distribution. The corresponding likelihood function is given by:

$$L(\lambda_i) = \prod_{i=1}^{Y} \frac{\Gamma(\theta + y_i)}{\Gamma(\theta) y_i!} \left[\frac{\theta}{\theta + \lambda_i}\right]^{\theta} \left[\frac{\lambda_i}{\theta + \lambda_i}\right]^{y_i} \qquad 2.9$$

where Y is the total number of sites and all other symbols have already defined meanings. The likelihood function is maximized to obtain coefficient estimates for β and α.

The Negative Binomial model form used in this research has already been established in literature (Usman *et al*., 2011). The functional form is shown in the equation below.

$$\mu_i = (\exp osure)^{\beta_1} * \exp(\beta_0 + \beta_2 x_{i1} + \beta_3 x_{i2} + \ldots\ldots\ldots + \beta_k x_{ij}) \qquad 2.10$$

Where μ_i is the predicted number of collisions (response variable), $x_{i1}, x_{i2}, x_{i3}, \ldots\ldots x_{ij}$ are the predictors (explanatory variables), and $\beta_0, \beta_1, \beta_2, \beta_3, \ldots\ldots \beta_k$ are the parameters (coefficient) estimates from the regression model.

2.5 Chapter Summary

Various research work in the area of collision and crime prediction were discussed, and, in an attempt to predict the frequency of collisions, linear regression modeling technique has been used by some researchers, using the relationship between collisions and traffic volumes (Ceder and Livneh, 1982; Ceder, 1982). However, to apply linear regression model as a modelling technique, some assumptions must be valid: variables must be normally distributed; a linear relationship should exist between independent and dependent variables; there is no auto-correlation among variables, there is no or little multicollinearity; and the assumption of homoscedasticity, which implies that the errors or residuals associated with a model are unequal across the range of the dependent variable (laffont *et al*., 2004). If these assumptions are not met, the results from such a model will be either over-estimated or under-estimated.

Jovanis and Chang (1986) argued that, according to Cresswell and Froggatt (1983), it was common to think traffic collisions follow a Poisson or Bernoulli process, which implies that the variance of collision frequency is directly related to its mean. A Poisson process suggests a continuous time of arrival or occurrence of events whereas a Bernoulli process assumes a discrete arrival or occurrence of events. Also, Poisson is an exponential function whereas a Bernoulli process is a geometric function. Both processes imply an increase in collision frequency is a result of increase in its mean because, in a Poisson distribution error structure, variance is equal to the mean. Higher collisions are expected at locations with high traffic volumes because of increase in conflicts. Therefore, an increase in traffic volume, which is one of the most important explanatory variables in any collision prediction model, implies an increase in mean and variance. This increase, therefore, downplays the assumption of homoscedasticity, affects the confidence levels of estimates of the model parameters, and invalidates any hypothesis tests concerning the significance of the parameters. Moreover, the non-negativity of collisions imposes a restriction on the use of linear regression in collision prediction models. Therefore, a Poisson or Negative Binomial distributions are the most appropriate techniques to model collisions.

Furthermore, criminal activities as events have similar characteristics to collisions. Because crimes are random, rare, and non-negative, Negative Binomial has been used by various researchers in predicting crime. Evidently, Negative Binomial presents the most accepted statistical modeling technique applied to aggregated data. The unit of analysis for aggregated or macro-level study should be carefully chosen to prevent bias, which has the potential to compromise the outcome of the analysis. If the intended purpose is for enforcement or implementation of a countermeasure, micro-level analysis is recommended, but if the sole aim is for planning purposes, macro-level analysis is preferred. However, a joint micro and macro level model can be adopted to achieve both purposes.

CHAPTER 3 . DATA

This chapter is divided into four sections and describes the data collection process and sources as well as the study area for this research. Statistical descriptions of the data collected are also presented. The first section describes the study area for which spatial analysis for this research is conducted. Geography and other information of the City of Regina are presented in this section as well. The second section of this chapter focuses on data sources as well as the data used in collision and crime prediction models. The issues of boundary data assignment in data aggregation are addressed in the third section. The last section of this chapter presents descriptive and spatial data analysis for the various data collected.

3.1 Study Area

The City of Regina, the provincial capital, is the second largest city in Saskatchewan with an estimated population of 241,400 (Statistics Canada, 2015). With the fourth highest population growth rate of 3.085% in Canada, the growth rate is mostly due to immigration and births (Leader Post, 2014). The City of Regina has four provincial highways as well as the Trans-Canada Highway 1. The study area, the City of Regina's geographical limits, has been divided into 299 Traffic Analysis Zones (TAZs), and the map in Figure 3.1 shows the city limits. These TAZs will be used as the unit of analysis for both traffic collisions and crimes. Various data collected were assigned to the individual TAZ based on their geographical location or coordinates. All maps in this thesis were developed using the North American Datum 1983, Universal Transverse Mercator Zone 13 North (NAD83, UTM Zone 13N). The geographic coordinate system used is the geographic coordinate system (GCS) North American 1983 with Greenwich being the prime meridian. Each TAZ already has demographic, some land use, and economic information embedded in them as a GIS-ready file. Traffic Analysis Zones are already determined areas by the City for planning and transportation planning purposes. The information embedded in the GIS-ready file included population information, housing density information, parking cost information, graduate student enrolment, and land use. These variables will be further described in this thesis. In the data assignment stage of this research, we discovered some Traffic Analysis Zones had no traffic volume data, signifying these were either open space, rivers, or undeveloped land areas with no road networks; those zones were excluded from analysis. As such, 262 Traffic Analysis Zones were used in the modeling stage of this research.

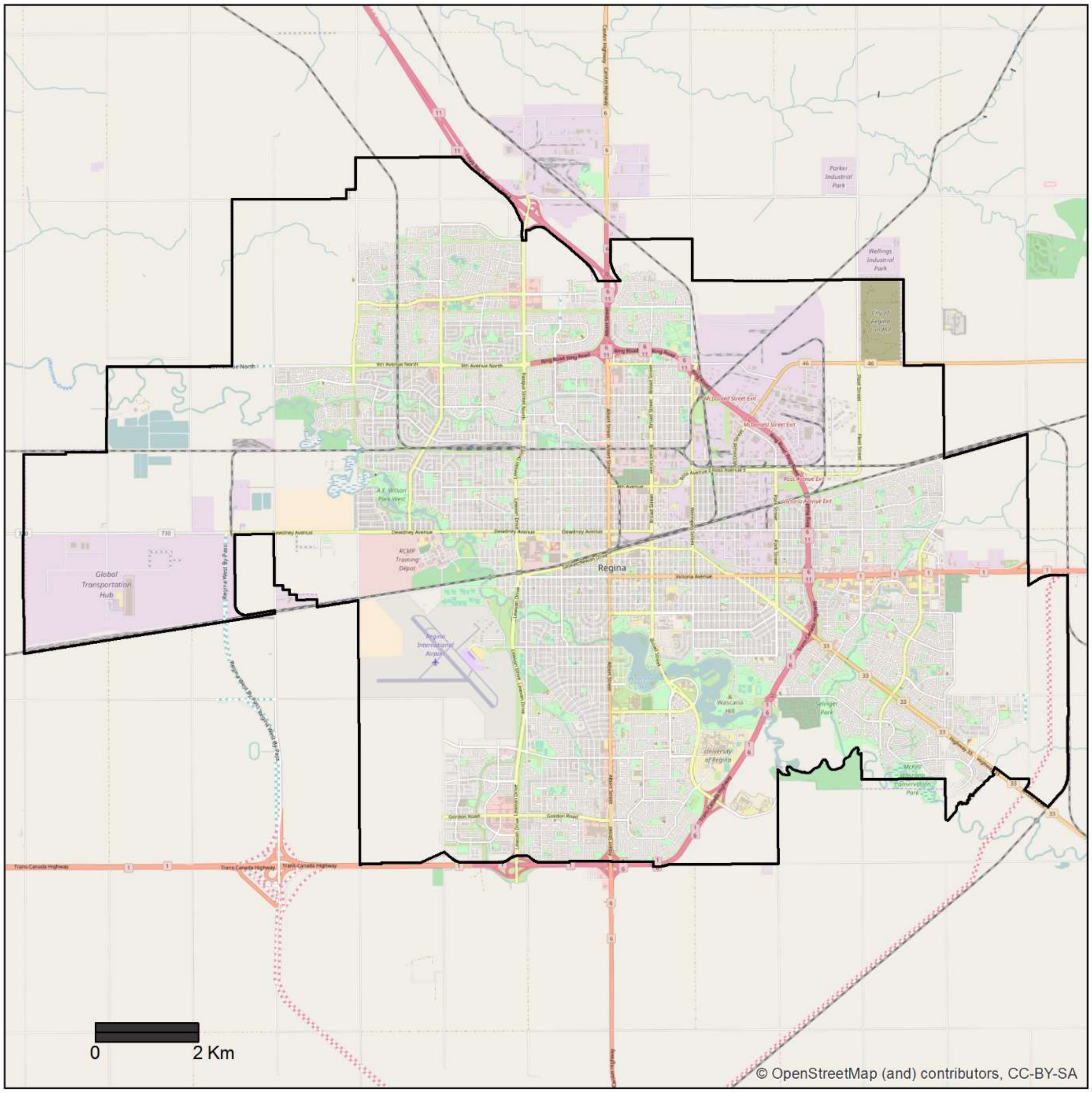


Figure 3.1: Map Showing the City of Regina City Limits

3.2 Data Source

Data used for this research were obtained from three agencies and an open data source. These agencies included the following: Saskatchewan Government Insurance, Regina Police Service, and the City of Regina. Traffic Analysis Zone map data were obtained from the OpenData website by the City of Regina. Data collected included sociodemographic information, traffic collision data, roadway infrastructure data, land use data, crime data, and City of Regina base maps.

3.2.1 Traffic Collision Data

Traffic collision data acquired from Saskatchewan Government Insurance came in three separate formats: accident table, vehicle table, and occupant table. Each set contained specific information related to the three main formats. The Accident table contains information relating to the general circumstances of the collision, such as the date, time, location, severity, number of people killed or injured, posted speed limit, location identifier known as urban grid (UGRID), collision cost, and weather information. UGRID are designated points in a road network with known x and y coordinates used to record information. Assigned to intersections and road segments, UGRID are very unique to each intersection and a specified stretch of road segment. For instance, on a 2 km stretch of a road segment, each 500 m of road is assigned a unique UGRID. Information about the vehicles and drivers involved in a collision, such as the driver's age, gender, and date of birth, as well as the vehicle number and major contributing factors to collisions, are contained in the vehicle table. Finally, the occupant table contains detailed information about each person involved in each injury collision. In all tables is a common field referred to as the case number, which uniquely identifies each traffic collision. Various information needed for the analysis was extracted from these three tables and combined for further analysis.

Five-year (2009-2013) collision data were collected for analysis, and three severity levels were considered for modeling: Property Damage Only, Fatal-Injury, and Total number of collisions. Property Damage Only collisions are traffic collisions in which no person sustains an injury, but there is damage to private or public property that costs more than $1000 (Capital Region Intersection Safety Partnership, 2012). Property Damage Only usually collisions result in deformity or destruction of parts of vehicles involved in the collision and/or public property, such as medians, electric poles, and installed signage along road segments. Injury collisions, on the other hand, result in at least one person sustaining an injury but not leading to death (Capital Region Intersection Safety Partnership, 2012). Fatal collisions result in at least one person sustaining bodily injury and resulting in death within 30 days of the date of the collision (Capital Region Intersection Safety Partnership, 2012). Fatal and injury collisions were combined and referred to as fatal-injury collisions. Fatal-Injury collisions are of major concern to transportation engineers. Total collisions are the aggregate of all the different levels of collisions. Over the study period, there were 29,411 recorded traffic collisions: 79.4% (23,366) were Property Damage Only collisions, and the remaining 20.6% (6010) were Fatal-Injury collisions.

Each collision record has a unique geolocation identifier known as UGRID. Based on the UGRID assigned to collision records, collisions were further assigned to the various Traffic Analysis Zones. In the data assignment process, there were issues. One main concern was boundary data assignment, and the other concern was collision data with unavailable UGRID; consequently, some data were lost in the process of aggregation. This issue will be further discussed in a later section of this chapter. Table 3.1 is a summary of traffic collisions recorded over the study period by severity and years.

Table 3.1: Collisions by Year and Severity

Year	Property Damage Only	Injury	Fatal	Fatal-Injury	Total
2009	5,270	1237	6	1243	6,513
2010	4,451	1169	8	1177	5,628
2011	4,376	1158	5	1163	5,539
2012	4,627	1309	8	1317	5,944
2013	4,642	1137	8	1145	5,787

Figures 3.2, 3.3, and 3.4 are histograms depicting TAZ-level collisions frequency distribution for Fatal-Injury, Property Damage Only, and Total collisions, respectively. Labels on the y-axis indicate the number of Traffic Analysis Zones with the specified collision frequency on the x-axis.

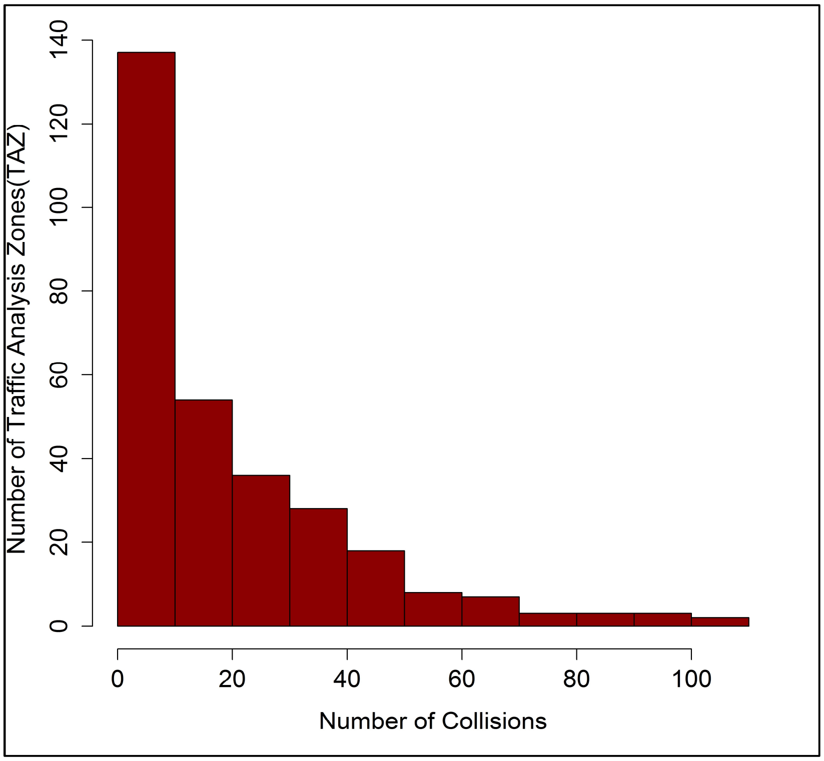


Figure 3.2: Frequency Distribution of Fatal-Injury Collisions

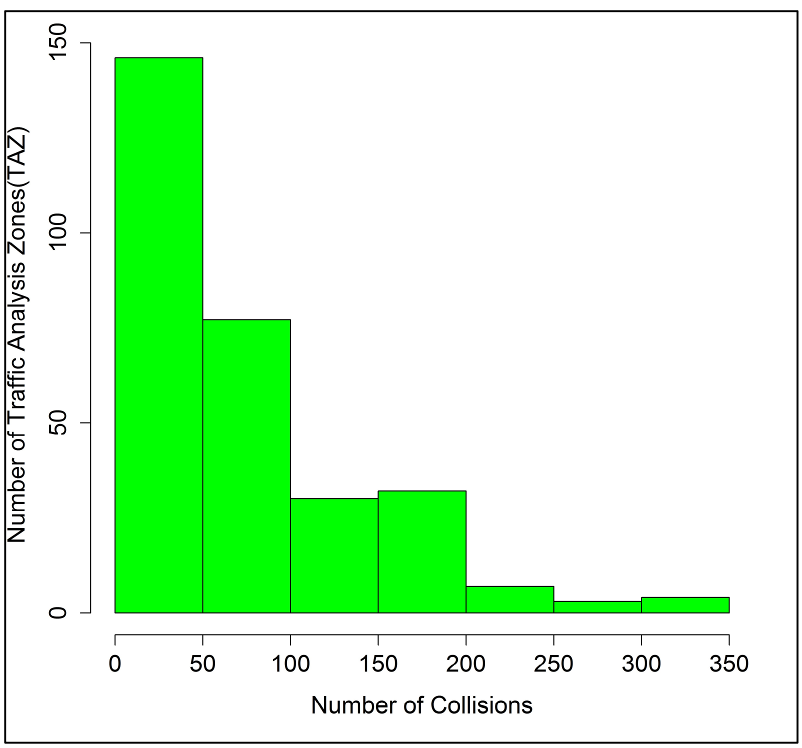


Figure 3.3: Frequency Distribution of Property Damage Only Collisions

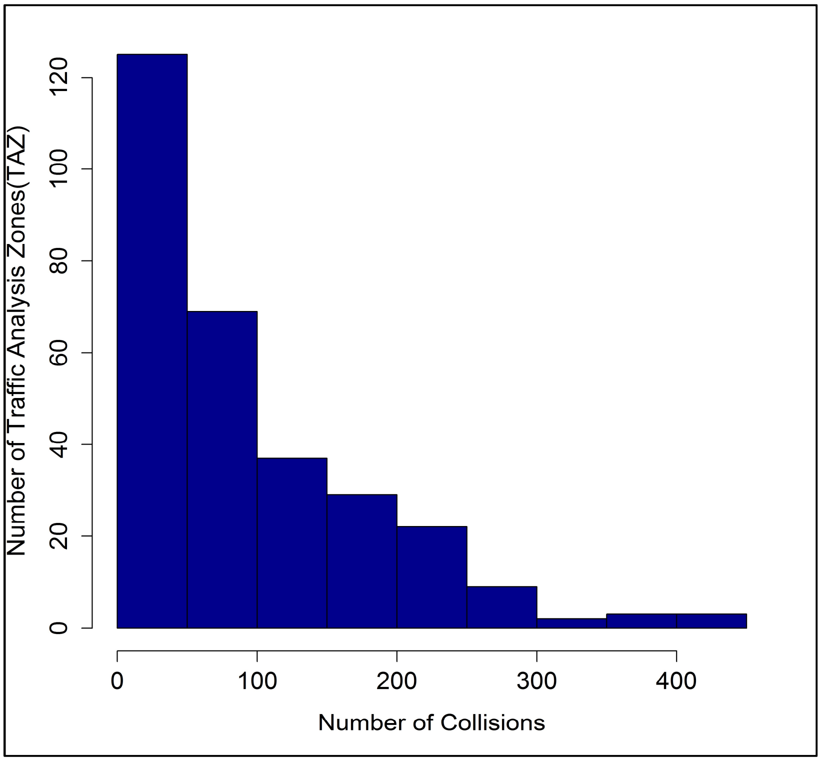


Figure 3.4: Frequency Distribution of Total Collisions

As an evident characteristic of count data, shown in Figures 3.2, 3.3, and 3.4 are skewed to the left, which implies negative binomial will be an appropriate technique to model collision data, because, Negative Binomial distribution traditionally has a distribution that is skewed to the left.

3.2.2 Traffic Volume

Traffic volume is the most important variable in any collision prediction model and is the main means for measuring exposure. The primary exposure variable in transportation planning and traffic engineering is the Annual Average Daily Traffic. Annual Average Daily Traffic is the total traffic volume passing an intersection or on a road segment in a year divided by 365 days. Average Annual Daily Traffic is obtained by performing traffic volume counts, conducted by either intrusive means, such as the use of pneumatic road tubes, piezoelectric sensors, and inductive loops, or non-intrusive means, such as manual counting, microwave radars, video image processing, and ultrasonic means. Because it is very expensive to perform these traffic volume studies, most cities do not conduct such studies for the entire city road network. Cities instead collect for limited portions, and those volumes are assumed to be the same for other road corridors of the same classification, namely arterial, collector, and local roads.

Therefore, 2014 Average Annual Daily Traffic data collected from the City of Regina did not have traffic volumes for all local roads. Accordingly, traffic volumes for local roads with missing volumes were assumed using values from other local roads with similar characteristics (Ni *et al*., 2005). This assumption is a popular practice in transportation planning and traffic engineering studies; however, in macro-level studies, Average Annual Daily Traffic cannot be used in analysis. Consequently, another variable, derived by multiplying Average Annual Daily Traffic by the road length, is calculated: Vehicle-Kilometer-Traveled. After assigning all missing Average Annual Daily Traffic volumes, it then became easier to calculate Vehicle-Kilometer-Traveled for all road segments. The arithmetic summation of all Vehicle-Kilometer-Traveled for roads within a Traffic Analysis Zone are then assigned to that particular Traffic Analysis Zone. This aggregated Vehicle-Kilometer-Traveled then becomes the main exposure variable.

3.2.3 Crime Data

Crime data, acquired from the Regina Police Service, came in three parts: occurrence file, address file, and person file. All three parts had a common field, known as occurrence file number, and each crime incident had a unique occurrence file number. Each occurrence file contained information about the times the crimes occurred: start time, end time, and reported time. Every address file included information about the location of each crime, including the district, street name, and x and y coordinates. Lastly, the person file contained information about persons involved in each crime, including year of birth, gender, citizenship, and birth country. During the study period for this research, there were over 90,000 reported crimes but 65,505 were used in this research. Because there were many crime occurrence types, crimes were grouped into 10 general classes in consultation with the Regina Police Service. Table 3.2 summarizes the various crime occurrence types and the general classes into which they were grouped.

Table 3.2: General Crime Groupings (Regina Police Service Classification)

General Class	Crime Occurrence Type
Assault	Aggravated Assault-Level 3 Assault-Common Level 1 Assault-Other CC-Criminal Negligence causing bodily harm Assault with weapon or cause bodily Harm Level 2 Discharge Firearm with intent DVC Aggravated Assault-Level 3 DVC Assault-Common Level 1 DVC Assault with weapon or cause bodily Harm-Level 2 Pointing a firearm Using firearms (or imitation) in commission of offence
Arson	Arson Arson-Disregard for Human Life
Break and Enter	Break and Enter Break and Enter-Home Invasion Break and Enter-Firearms Break and Enter-Compound Break and Enter to Motor Vehicle-Firearms DVC Break and Enter
Robbery	Robbery Robbery-Commercial Robbery-Delivery Person Robbery-Financial Institution Robbery-Purse-snatching Robbery-Street Robbery-Taxi
Sexual Assault	Aggravated Sexual Assault DVC Aggravated Sexual Assault Sexual Assault Sexual Assault with a weapon DVC Sexual Assault DVC Sexual Assault with a weapon Incest Invitation to sexual touching Luring a child via a computer Sexual exploitation Sexual exploitation of a person with a disability Sexual interference Sexually explicit material available to a child Voyeurism
Theft	DVC Theft Over $5,000 DVC Theft Under $5,000 Mail theft before delivery over $5,000

*DVC- Domestic Violence Court

Table 3.2: General Crime Groupings (Regina Police Service Classification) [cont'd]

Theft	Mail theft before delivery under $5,000 Pick-pocketing over $5,000 Pick-pocketing under $5,000 Purse-snatching under $5,000 Shoplifting Over $5,000 Shoplifting $5,000 or under Theft Over $5,000 Theft Under $5,000 Theft of telecommunications over $5,000 Theft of telecommunications under $5,000
Mischief	Mischief-No damage Mischief Over $5,001 Mischief Under $5,000- Graffiti Mischief Under $5,001 Mischief Willful Damage Public Mischief DVC Mischief Over $5,001 DVC Mischief Under $5,001
Theft from Auto	Theft from Auto Over $5,000 Theft from Auto Under $5,000
Theft of Auto	Theft of Auto Theft of Auto Over $5,000 Theft of Auto Under $5,000

*DVC- Domestic Violence Court

The definition of each individual crime type will not be discussed since it is beyond the scope of this research. The chart below shows the distribution of crime types by general classes. Figure 3.5 is a pie chart illustrating the proportion of each type of crime compared to the total number of crimes, and indicated in brackets are the actual numbers of each type of crime.

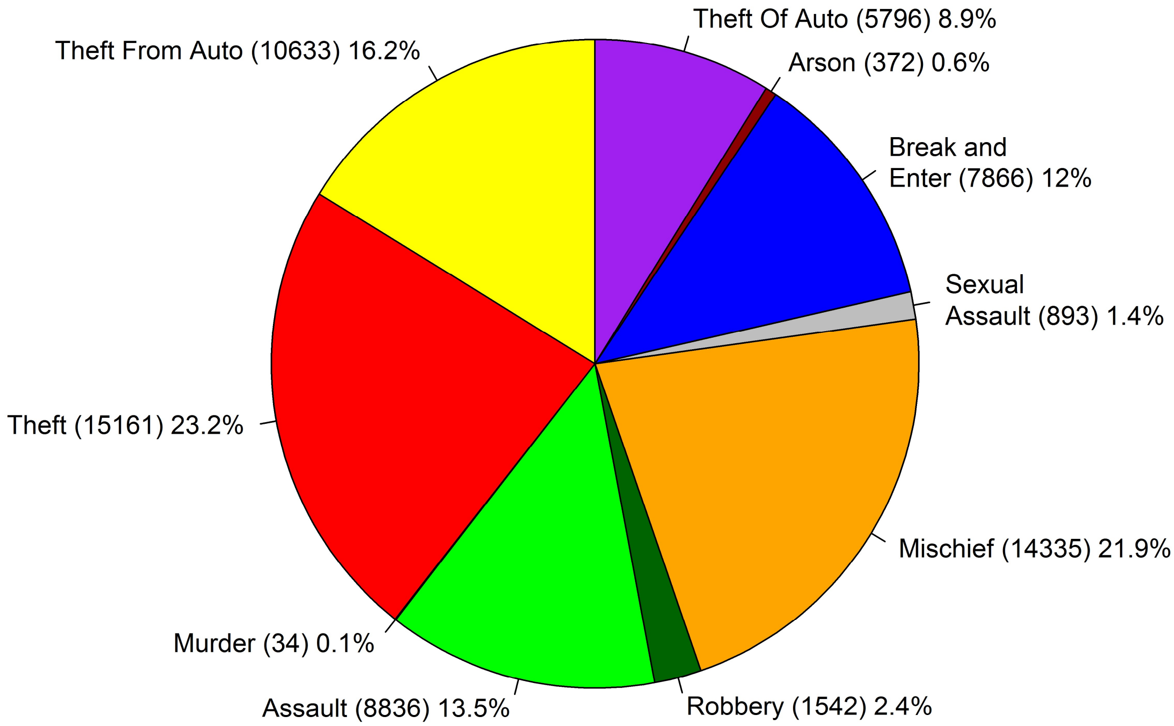


Figure 3.5: Total number and percentage of different types of crimes (2009-2013)

The ten classes of crimes were further grouped into two types: violent and non-violent crimes. Violent crimes constituted assault, sexual assault, murder, robbery, and arson. Non-violent crimes include theft, theft from auto, theft of auto, break and enter, and mischief. One major issue with the crime data had to do with the time of occurrence of crimes. As previously mentioned, each crime has three columns that had time information: start, end, and reported. However, some crimes had no recorded start, end time, nor reported time in conducting exploratory statistics to determine trends of crime. Nonetheless, every crime incident had reported times, as such, the reported time was used as the reference time for performing time-trend analysis.

Figure 3.6, 3.7, and 3.8 are histograms depicting the distribution of TAZ-level non-violent, violent and total crimes data respectively. The Negative Binomial model distribution function is skewed to the left similar to the ones in Figures 3.6, 3.7 and 3.8 therefore, Negative Binomial is an appropriate modeling technique for the aggregated crime data.

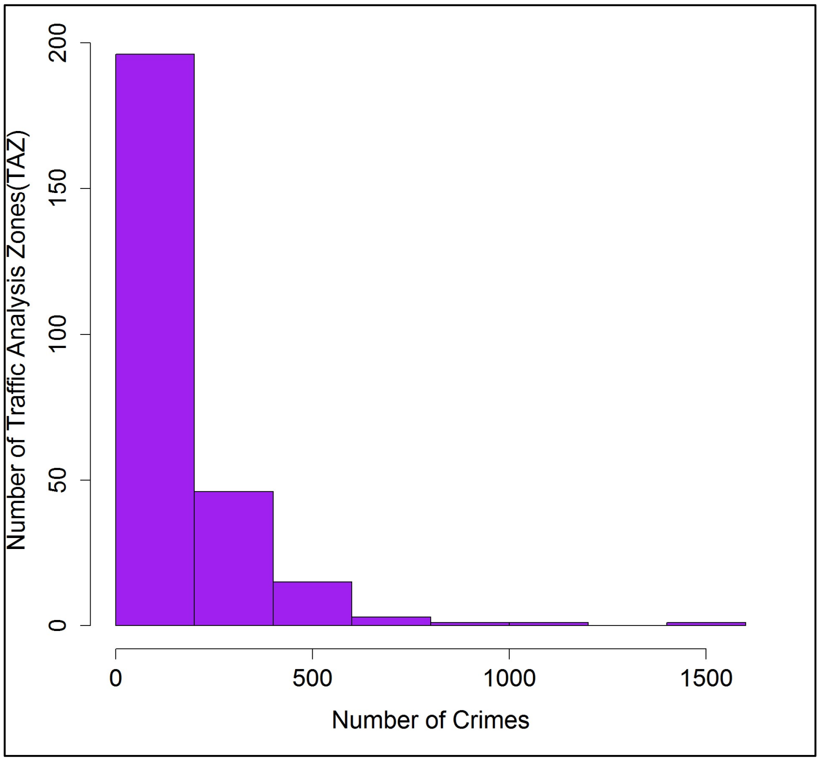


Figure 3.6: Frequency Distribution of Non-Violent Crimes

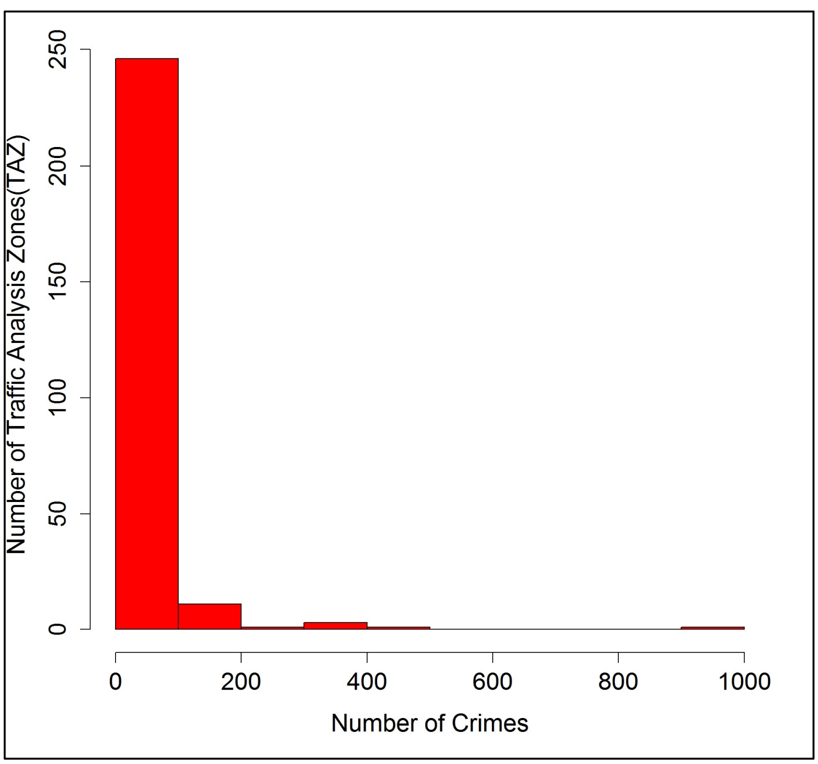


Figure 3.7: Frequency Distribution of Violent Crimes

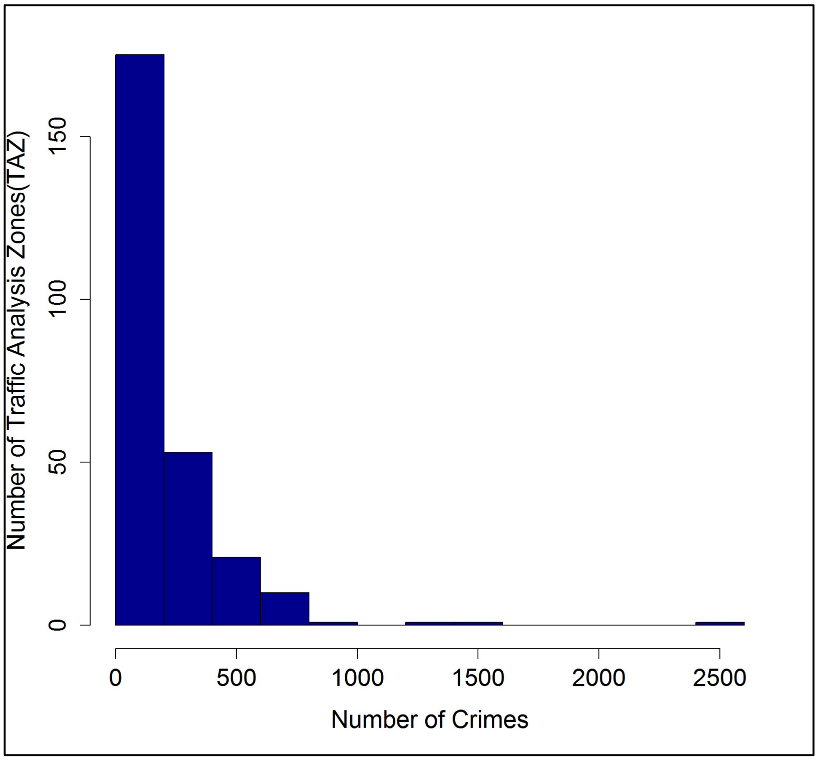


Figure 3.8: Frequency Distribution of Total Crimes

3.2.4 Socio-Economic and Demographic

Socio-economic and demographic data (2012) were obtained from the City of Regina. These data were already assigned to Traffic Analysis Zone by a unique identifier for each zone, referred to as Traffic Analysis Zone number. Tables 3.3 and 3.4 present the description of demographic and socio-economic data collected from the City of Regina, respectively.

Table 3.3: Demographic variables and description

Variable	Description (per Traffic Analysis Zone)
POP_01to17	Population of residents aged 1 to 17
POP_18to24	Population of residents aged 18 to 24
POP_25to44	Population of residents aged 25 to 44
POP_45to64	Population of residents aged 45 to 64
POP_65plus	Population of residents aged 65 and above
TOT_POP	Total Population of residents
POP_DENSITY	Population Density of residents (Population/Sq. km.)
NO_GRDSCH	Number of residents Enrolled in Graduate School
NO_PSSTD	Number of residents Enrolled in a Post-Secondary Institution

Table 3.4: Socio-economic variables and description

Variable	**Description (per Traffic Analysis Zone)**
OFFICE_AREA	Total land area allocated for office use (m^2)
RETAIL_AREA	Total land area allocated for retail use (m^2)
INDUSTRY_AREA	Total land area allocated as industry use (m^2)
HOSPT_SPACE	Total land area allocated as hospital space use (m^2)

3.2.5 Road Network

The City of Regina road basemap, GIS-based shapefile, was collected from the City of Regina. The shapefile came in two forms: road segments and intersections. These two files had several exposure variables about the entire city road network, as well as road infrastructure. Because they came in GIS-ready format, it was easy to assign those data to the various Traffic Analysis Zones by geolocation. Table 3.5 is a summary of the description of exposure variables. In addition to Vehicle-Kilometer-Traveled, these exposure variables will be very important variables to be considered at the modeling stage of this research.

Other variables were also derived from the existing data obtained, and Table 3.6 is a description of those derived variables.

Table 3.5: Exposure Variables and descriptions

Variable	Description (per Traffic Analysis Zone)
AVE_SPDLIM	Average Speed Limit, km/hr
TOT_ROADLEN_20	Total Road length with posted speed 20km/hr
TOT_ROADLEN_30	Total Road length with posted speed 30km/hr
TOT_ROADLEN_40	Total Road length with posted speed 40km/hr
TOT_ROADLEN_50	Total Road length with posted speed 50km/hr
TOT_ROADLEN_60	Total Road length with posted speed 60km/hr
TOT_ROADLEN_70	Total Road length with posted speed 70km/hr
TOT_ROADLEN_80	Total Road length with posted speed 80km/hr
TOT_ROADLEN_100	Total Road length with posted speed 100km/hr
TOT_ROADLEN	Total Road length, m
AVE_ROADLEN	Average Road Length, m
NO_ROADS_PER_TAZ	Number of Road Segments
NO_3LEGS_INT	Number of Three-Leg Intersections
NO_4LEGS_INT	Number of Four-Leg Intersections
NO_5LEGS_INT	Number of Five-Leg Intersections
TOTAL_INT	Total Number of Intersections
INT_DEN	Intersection Density defined as the number of Intersections divided by Traffic Analysis Zone Area (intersections/ sq. m.)
ARTERIAL_LEN	Total Arterial Road Length, m
COLLECTOR_LEN	Total Collector Road Length, m
DRIVEWAY_LEN	Total Driveway Length, m
EXPRESSWAY_LEN	Total Expressway Length, m
HIGHWAY_LEN	Total Highway Length, m
LOCAL_ROAD_LEN	Total Local Road Length, m
PRIVATE_ROAD_LEN	Total Private Road Length, m
RAMP_LEN	Total Ramp Length, m
ROW_LEN	Total Right of way Length, m

Table 3.6: Derived Road Network Variables

Variable	Description (per Traffic Analysis Zone)
Total Lane Kilometer (TLKM)	The sum of all road segments in a Traffic Analysis Zone with units of kilometer
Vehicle-Kilometer-Traveled (VKMT)	Exposure variable obtained by multiplying Average Annual Daily Traffic by total-lane-kilometer
Three-Leg Intersection Proportion (I3WP)	Proportion of the number of 3-leg intersections as a ratio of total number of intersections
Arterial Road Length Proportion (ALKP)	Proportion of arterial road length as a ratio of the total-lane-kilometer
Local Road Length Proportion (LLKP)	Proportion of local road length as a ratio of the total-lane-kilometer

3.2.6 Traffic Analysis Zone and Land Use Data

Traffic Analysis Zone data (2015) which came with some land use data, was also acquired from the City of Regina open source data website. Different land use types, Traffic Analysis Zone perimeter, and Traffic Analysis Zone area data were collected from the City of Regina and added to the Traffic Analysis Zone map data in ArcGIS. Table 3.7 is a summary of Traffic Analysis Zone and land use data acquired and their descriptions. As assigned by city planners, some zones have multiple land uses, and that data have been captured by the number of land uses. Figure 3.9 is a map, showing all Traffic Analysis Zones that were used for prediction models.

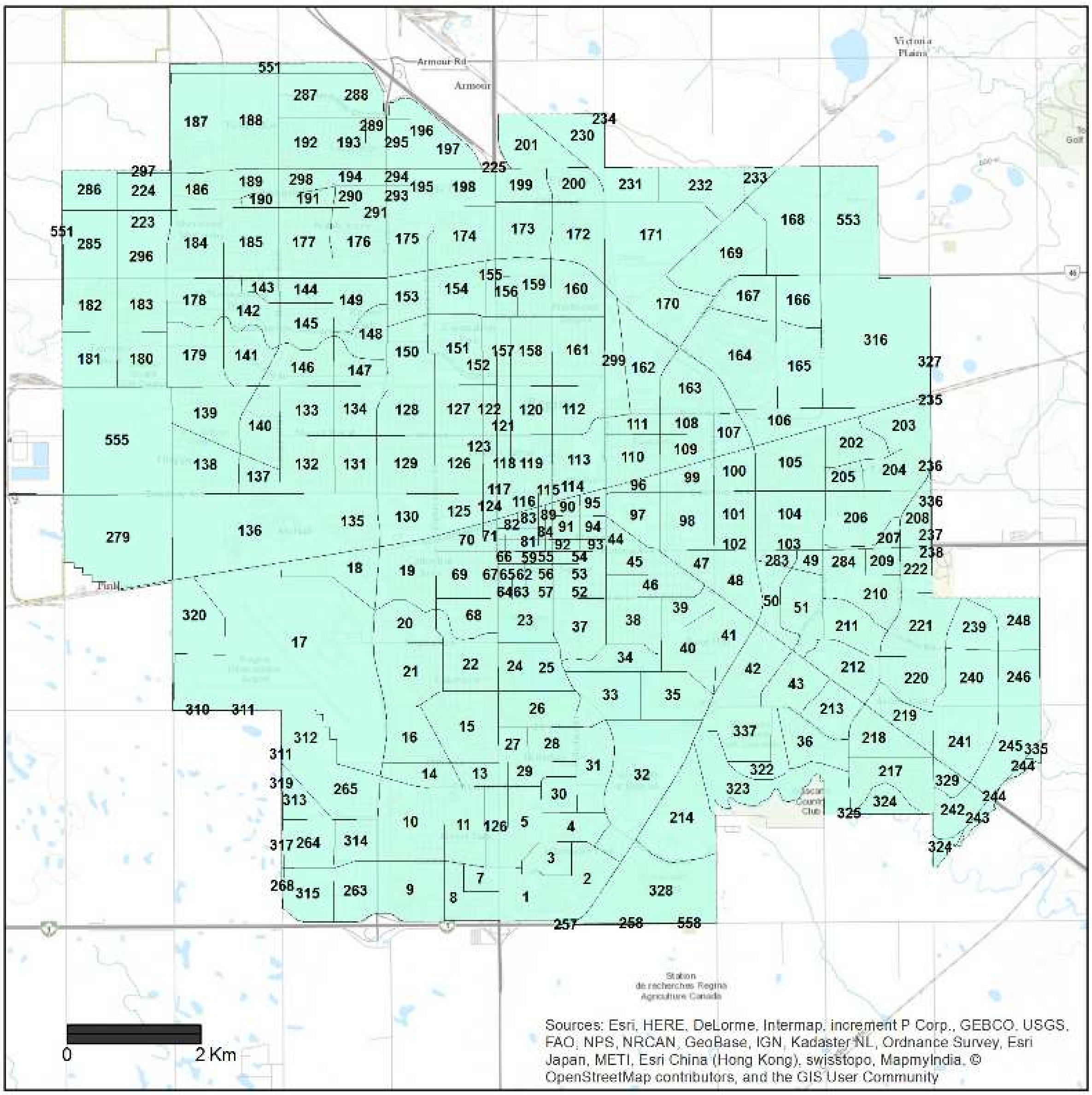


Figure 3.9: Traffic Analysis Zone used in prediction models

Table 3.7: Traffic Analysis Zone and Land Use Variables Description

Variable	Description
AIRPORT_AREA	Total Area designated for Airport operational use and controlled by Transport Canada under *The Aeronautics Act* (Canada), m^2
COMMERCIAL_AREA	Area intended to provide commercial, personal service, business development, office businesses and retail services, m^2
INDUSTRIAL_AREA	Area intended to provide industrial uses engaged in manufacturing, processing, assembly, distribution, service and repair activities, m^2
INSTITUTIONAL_AREA	Area designated to provide space for public owned facilities of an institution and community service, m^2
OFFICE_AREA	Area designated for development of office areas outside of downtown area to provide alternate market for businesses with close proximity to major corridors, regional customers, intermodal hubs, *etc*.
OPENSPACE_RECREATION_AREA	Total area designated for open space/recreation purposes, m^2
RAILWAY_AREA	Area designated for land use directly associated with provision of transportation by railroad, switching, and terminal operations, m^2
RESIDENTIAL_HD_AREA	Area designated to provide housing with a net density in excess of 50 dwelling units per hectare, m^2
RESIDENTIAL_LD_AREA	Land area designated to provide housing with a net density below 25 dwelling units per hectare, m^2
RESIDENTIAL_MD_AREA	Total Area Designated to provide for flexibility in building and site design with a net density of 25-50 dwelling units per hectare, m^2
URBAN_HOLDING_AREA	Area intended to protect lands required for future urban developments, m^2
NO_LU_PER_TAZ	Numbers of Different Land use per Traffic Analysis Zone
TAZ_PERIMETER	Perimeter of Traffic Analysis Zone, m
TAZ_AREA	Area of Traffic Analysis Zone, m^2

3.2.7 Integrated Database

All variables and response variables were then assigned to their respective Traffic Analysis Zone using Microsoft Excel, Microsoft Access, and ArcGIS. The aggregated database, which was then saved in ready-to-use format for R-language, which is the statistical software used for regression analysis as well goodness-of-fit tests. R was then used to randomly partition the integrated database for model calibration and validation. Model Calibration and validation will be discussed in Chapter Four.

3.2.8 Database Management and Problems

In developing collision data from the three sources, there was an issue of multiple road segments with duplicate UGRIDs. Each duplicate entry was checked manually and merged into a single segment when appropriate. Moreover, the road network basemap obtained from the City of Regina did not have information about the number of lanes for each road segment and whether a road segment was divided or not. This necessary information was gathered by checking Google Maps, as well as a city map, and manually assigning those attributes to the road segments.

3.3 Boundary Data Assignment

The boundary data assignment issue arises when collisions or crimes occur on road segments or intersections that serve as boundary for multiple Traffic Analysis Zones. This issue required time-consuming efforts to allocate collisions and crimes appropriately, to accurately represent the true collision and crime frequencies per each Traffic Analysis Zone. This process involves identifying collisions and crimes that occur at boundaries and further identifying the number of Traffic Analysis Zones that share that boundary. After identifying these Traffic Analysis Zones, numbers of collision and crime are assigned to these neighbouring boundary Traffic Analysis Zones. Figure 3.10 illustrates typical instances of boundary data among four, three, and two Traffic Analysis Zones. The first picture has an intersection, serving as boundary for 4 Traffic Analysis Zones. The second picture on the right illustrates another intersection at the boundary of 3 Traffic Analysis Zones. Lastly, the third picture represents a road segment that lies on the boundary of two Traffic Analysis Zones.

Figure 3.10: Typical Boundary Data Issues

Literature was reviewed to identify the various boundary data assignment methods. Ihssian (2014) completed extensive work on model prediction based on the boundary data assignment method among adjacent Traffic Analysis Zones. Some of the methods used are explained as follows:

- Equal-proportion-based: there is an equal split of boundary data between adjacent Traffic Analysis Zones. For instance, if data falls on a boundary between two Traffic Analysis Zones, the data is divided between the two Traffic Analysis Zones half and half (0.5:0.5). Data on a three adjacent Traffic Analysis Zones is split one-third to each Traffic Analysis Zone *etc*.
- Population-proportion-based: boundary data is split between adjacent Traffic Analysis Zones in proportion to their population. Traffic Analysis Zones with higher populations, are assigned higher numbers of collisions or crimes.

- Population-employment-based: boundary data is split between adjacent Traffic Analysis Zones in proportion to the summation of population and employment within each Traffic Analysis Zone.
- Total-lane-kilometer-based: boundary data is split based on proportion of total-lane-kilometer, Traffic Analysis Zones with higher Total Lane Kilometers, which is simply a much larger road network are assigned higher proportion of the boundary data.
- Multiple-count-based: in this approach, the same quantity of the boundary data is assigned to each adjacent Traffic Analysis Zone. For instance, if a Traffic Analysis Zone shares a boundary with two other Traffic Analysis Zones, and there are 20 collisions on that boundary, each Traffic Analysis Zone will be assigned 20 collisions. This method leads to double counting.

Ihssian's (2014) research revealed that the equal-proportion method proved most successful in terms of prediction. That approach divided the number of incidents (collision or crime, in this case) proportionally, based on the number of zones sharing the boundary or intersection. For instance, the boundary that shares two Traffic Analysis Zones (TAZ) will have 50% of the boundary data assigned to each Traffic Analysis Zone; for a three-TAZ boundary, one-third of the total data will be assigned to each of the three Traffic Analysis Zone; and, similarly, for a four-TAZ boundary, each zone will be assigned 25% of the data. Ihssian's (2014) approach was adopted in all boundary data issues in this current research. Because of the exclusion of some Traffic Analysis Zones from modeling, those zones were excluded before completing the data assignment process. Traffic collision and crime are count data and as such they are whole numbers. However, in situations where the boundary data is less than the number of boundaries, results in number of collisions and crimes having decimal numbers. for instance, one collision shared between two Traffic Analysis Zones results in each Traffic Analysis Zone having 0.5 collision.

3.4 Descriptive Statistics and Spatial Analysis of Data

3.4.1 Traffic Collision

Table 17 shows the proportion of data for the different observed collision severity types within City of Regina's Traffic Analysis Zones. In Table 3.8, observations refer to the number of Traffic Analysis Zones and as can be seen, 78% of assigned collisions are property damage only and 22% are fatal-injury collisions.

Table 3.8: Descriptive Statistics of collision data (2009-2013)

Variable	Observations	Total	Minimum	Maximum	Mean	Standard Deviation
Total Collisions	262	26,610	0	424	101.56	88.72
Property Damage Only Collisions	262	20,883	0	326	79.71	69.04
Fatal-Injury Collisions	262	5,759	0	103	21.98	21.65

Figure 3.11 shows the distribution of the observed numbers of total collisions per Traffic Analysis Zone. The number of observed total collisions were ranked: Traffic Analysis Zones with the top 10 highest numbers of observed number of collisions were assigned as high on the map; Traffic Analysis Zones with the top 11 to 20 number of collisions were assigned as medium; and the remaining locations were assigned as low. Table 3.9 provides further information about the range of numbers of collisions in each of the three classes; high, medium and low. Evidently, the map illustrates that a high number of collisions occur in Traffic Analysis Zone located in the downtown area and some residential areas.

Table 3.9: Observed Number of Total Collisions per Traffic Analysis Zone

Map Legend	Number of collisions	Number of Traffic Analysis Zones
High	285 - 424	10
Medium	284 - 242	10
Low	Less than 242	242

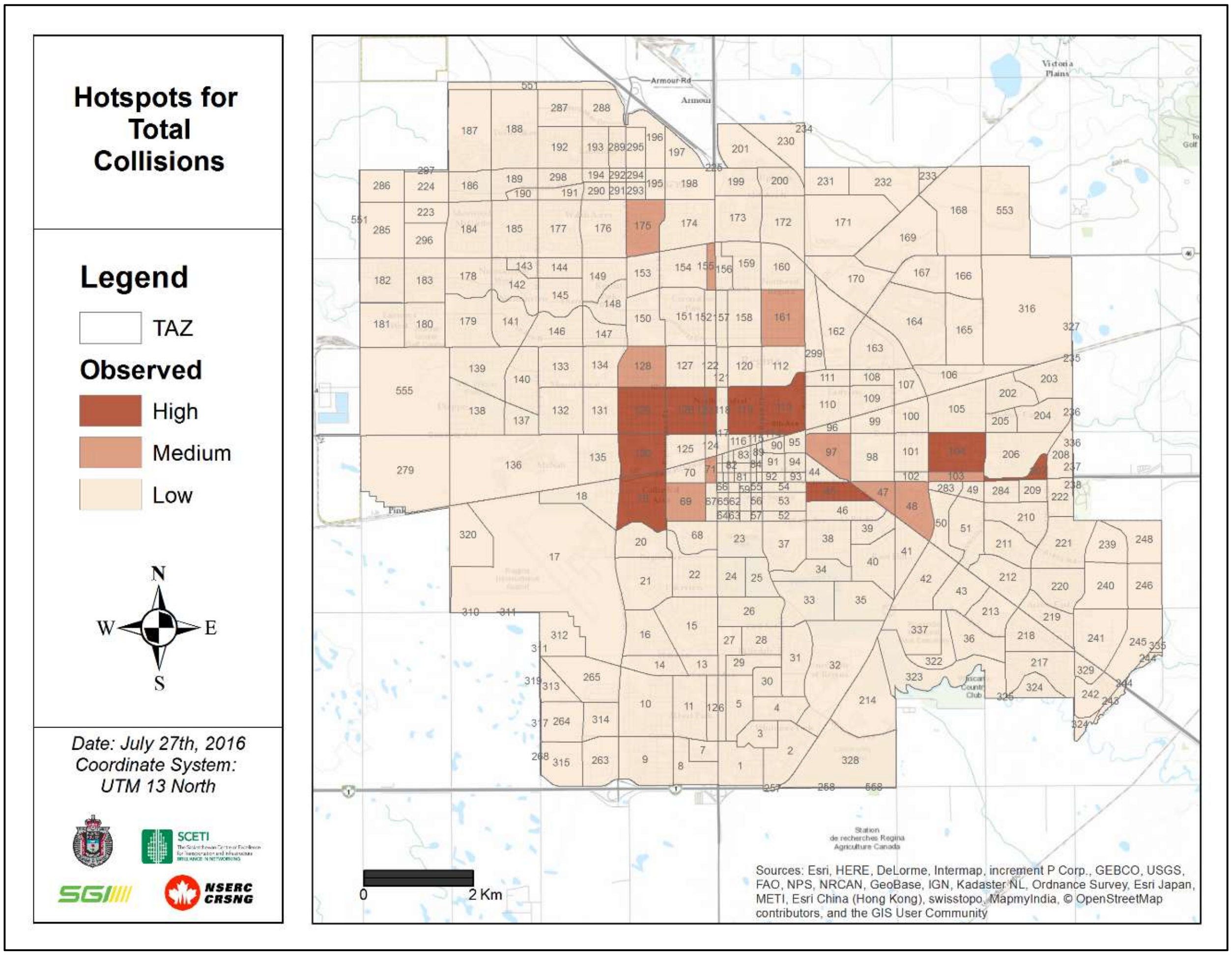


Figure 3.11: Map for Observed Number of Total Collisions Aggregated by Traffic Analysis Zone

Figure 3.12 presents a map depicting the distribution of the observed numbers of fatal-injury collisions per Traffic Analysis Zone. And Table 19 provides information about the range of values that have been labelled as high, medium, or low in the map in Figure 3.12. Fatal-Injury collisions seem to concentrate more in Traffic Analysis Zones located around the central business district and along high speed roadways. The observed numbers of Fatal-Injury collisions were ranked in the map: the top 10 were defined as high; the top 11 to 20 were assigned as medium; and locations ranked 21 and below were defined as low.

Table 3.10: Observed Number of Fatal-Injury Collisions per Traffic Analysis Zone

Map Legend	Number of collisions	Number of Traffic Analysis Zones
High	75 - 103	10
Medium	59 - 73	10
Low	Less than 73	242

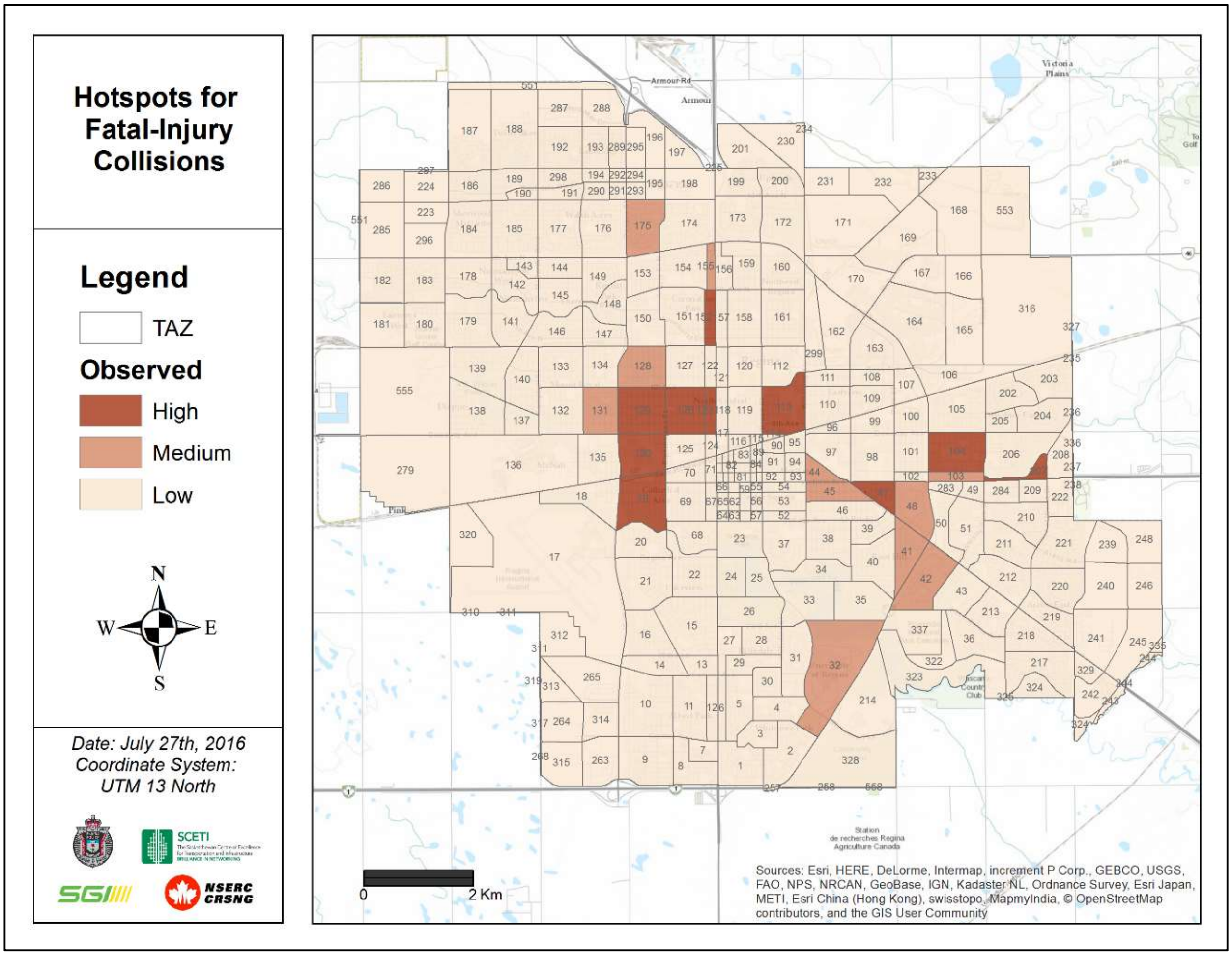


Figure 3.12: Observed Number of Fatal-Injury Collisions Aggregated by Traffic Analysis Zone

Figure 3.13 shows the distribution of the observed numbers of Property Damage Only collisions per Traffic Analysis Zone. Property Damage Only collisions are concentrated in the central business district of the City of Regina, as well as some residential areas not too far from the Central Business District. Table 3.11 includes the statistics about collision frequencies, representing high, medium, and low as indicated in the map, as well as the number of Traffic Analysis Zones that observed those frequencies.

Table 3.11: Observed Number of Property Damage Only Collisions per Traffic Analysis Zone

Map Legend	Number of collisions	Number of Traffic Analysis Zones
High	215 - 326	10
Medium	187 - 209	10
Low	Less than 187	242

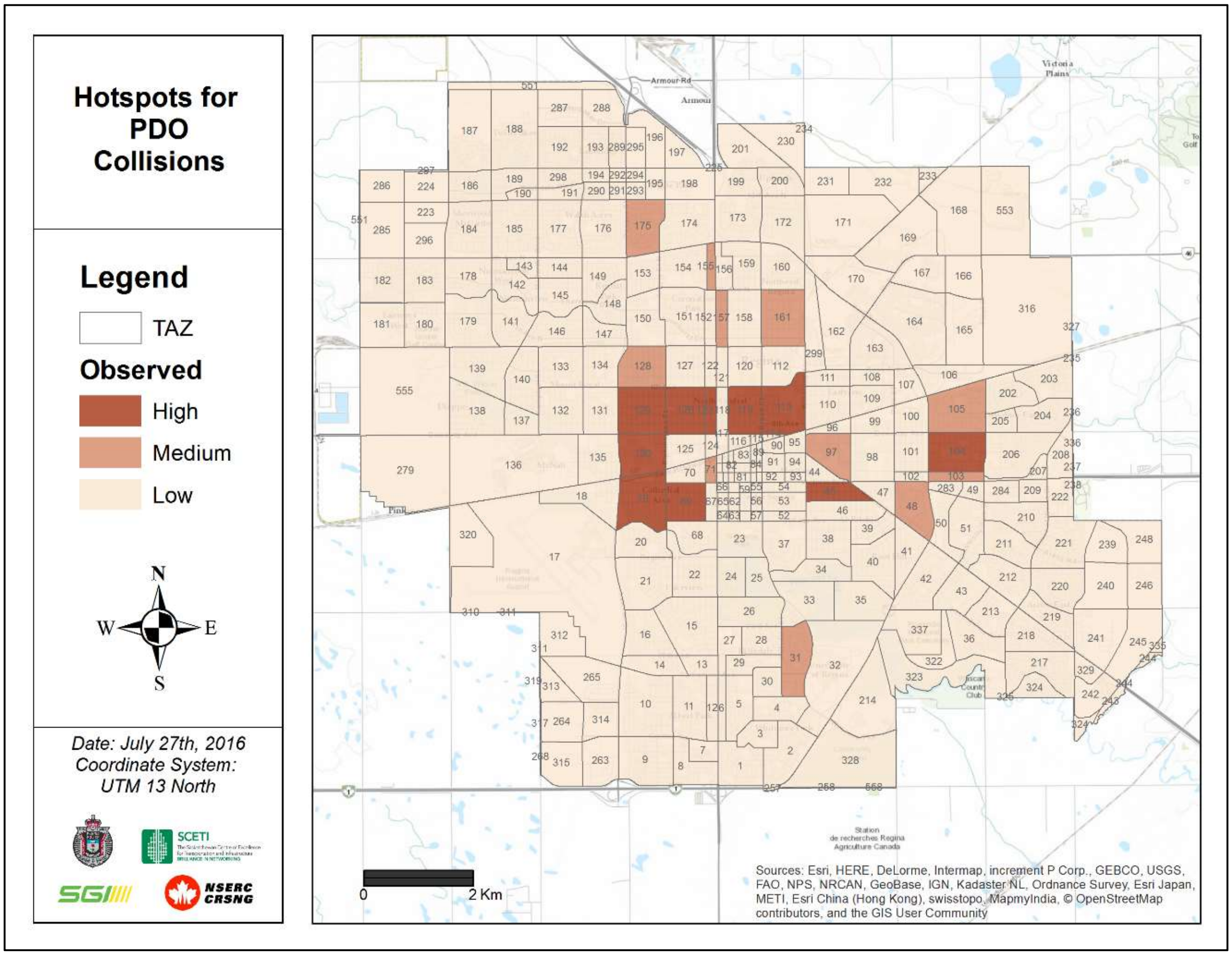


Figure 3.13: Observed Number of Property Damage Only Collisions Aggregated by Traffic Analysis Zone

3.4.2 Crime Data

Table 3.12 is a summary of descriptive statistics for the ten types of crime occurrences, and the bottom two refer to the two main groups of crimes: violent and non-violent crimes. The majority of the crimes were non-violent crimes, as can be seen from Table 3.12. Figure 3.14 is a map showing the spatial distribution of the observed number of total crimes aggregated by Traffic Analysis Zone. As expected in most cities, high numbers of crimes were recorded in the Central Business District area and nearby residential neighbourhoods. Traffic Analysis Zones were ranked by the observed number of total crimes, from highest to lowest. Traffic Analysis Zones with the top 10 frequency of observed number of total crimes are defined as high; locations with the top 11 to 20 numbers of total crimes are defined as medium; and locations ranked 21 and below are defined as low. Table 3.13 provides further information representing the range of values for each of the three levels of observed total crimes as labelled in the map in Figure 3.14; high, medium and low. As can be seen from Table 3.13, locations labelled as high has significantly high numbers of crimes over the study period (2009-2013) of this research, with values between 685 and 2422. These indicate areas that have experienced high numbers of crimes. Maps depicting observed numbers of crimes for the various classes and groups of crimes have been presented in Appendix C2.

Table 3.12: Aggregated Crime Data Statistics per Traffic Analysis Zone

Variable	Observations	Total	Minimum	Maximum	Mean	Standard Deviation
Assault Crimes	262	6,906.80	0	683.50	26.26	58.13
Break and Enter Crimes	262	6,057.60	0	290.5	23.03	32.27
Mischief Crimes	262	10,905.00	0	530.00	41.46	54.91
Robbery Crimes	262	1,300.60	0	141.00	4.95	12.18
Theft Crimes	262	11,421.20	0	593.20	43.43	77.78
Theft from Auto Crimes	262	8,037.60	0	174.80	30.56	30.67
Theft of Auto Crimes	262	4,648.50	0	241.00	17.57	26.62
Arson Crimes	262	268.70	0	60.30	1.02	4.06
Murder Crimes	262	24.00	0	7.00	0.09	0.52
Sexual Assault Crimes	262	609.00	0	66.00	2.32	6.20
Five-Violent Crimes*	262	18,231.80	0	9,115.90	69.06	563.37
Five Non-Violent Crimes*	262	53,794.20	0	12,747.00	203.77	794.08

*Two main groups of crimes: violent and non-violent

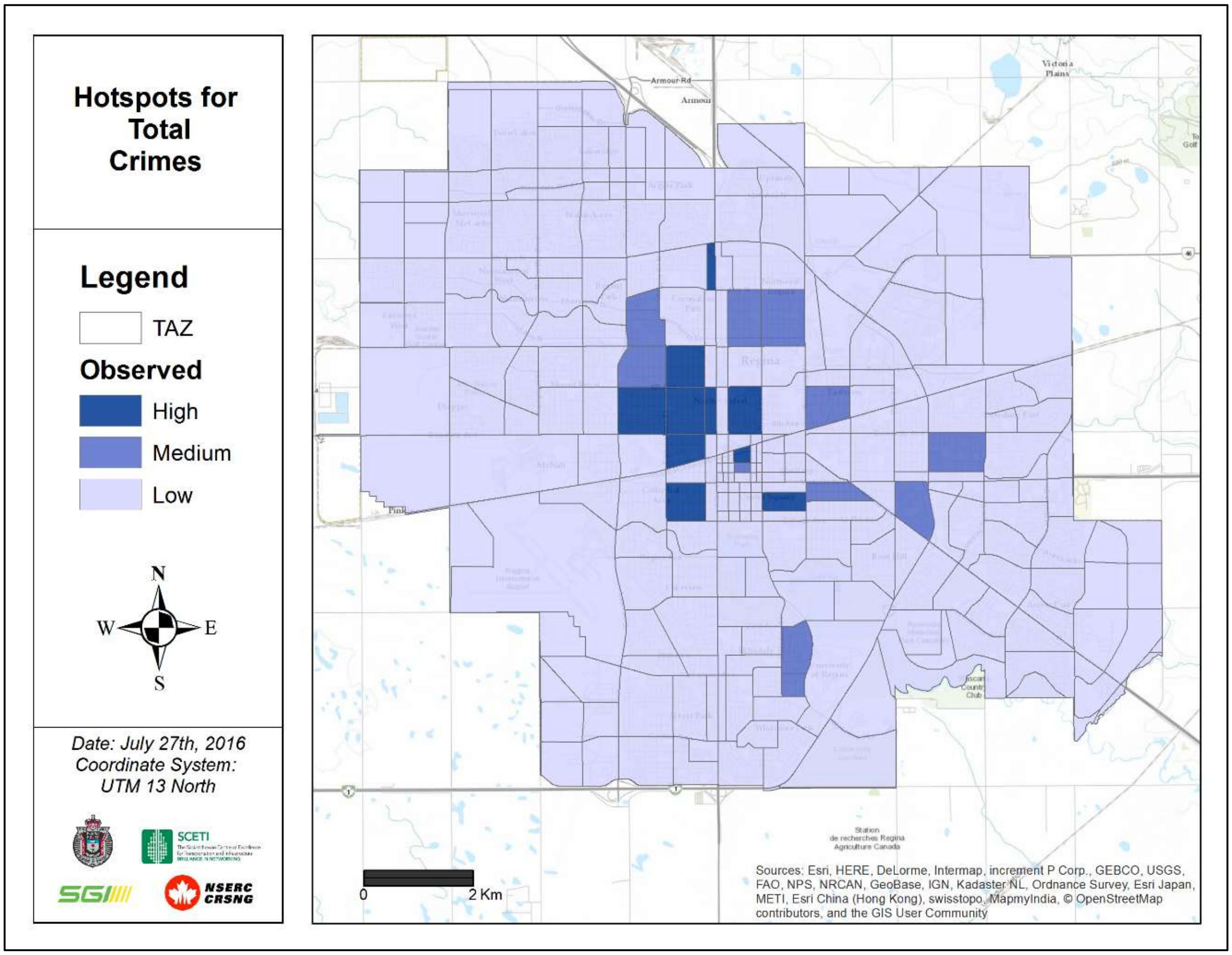


Figure 3.14: Map for Observed Number of Total Crimes

Table 3.13: Observed Total Crimes per Traffic Analysis Zone

Map Legend	Number of crimes	Number of Traffic Analysis Zones
High	685.1 - 2422	10
Medium	674.8 – 482.2	10
Low	Less than 674.8	242

3.4.3 Socio-Economic and Land Use Data

Statistics about socio-economic and land use data are presented in Table 3.14. Observations represents the number of Traffic Analysis Zones and total for each variable as well as the minimum, maximum, mean and the standard deviations are shown in Table 3.14.

Table 3.14: Aggregated TAZ-Level Socio-Economic and Land Use Statistics

Variable	Obs.	Total	Min.	Maximum	Mean	Standard Deviation
Office Space, m^2	262	902366.00	0	101970.00	3431.05	10266.25
Retail Space, m^2	262	1628686.00	0	122199.00	6192.72	13029.34
Industry Space, m^2	262	1613288.00	0	274205.00	6134.17	24540.63
Hospital Space, m^2	262	117231.00	0	66704.00	445.745	4732.10
Number of Land Use	262	1086.00	1	8.00	4.13	1.68
Commercial Area, m^2	262	7403575.00	0	312623.00	28150.48	46800.05
Institutional Area, m^2	262	4221201.88	0	620487.35	16050.20	46778.39
Open Space Recreational Area, m^2	262	24674256.21	0	2501580.25	93818.46	248948.94
Railway Area, m^2	262	2777417.00	0	691835.00	10560.52	52694.82
High Density Residential Area, m^2	262	4212293.98	0	196674.39	16016.33	32603.35
Low Density Residential Area, m^2	262	45363260.00	0	786855.00	172483.88	210194.97
Medium Density Residential Area, m^2	262	3664149.00	0	383555.00	13932.13	43650.73
Urban Holding Residential Area, m^2	262	21963243.00	0	3150675.91	83510.43	322005.57
Traffic Analysis Zone Area, m^2	262	137132121.34	0	6044326.67	521414.91	570457.13
Population Density, sq. km (km^2)	262	527664.94	0	10552.61	2006.34	1665.67
Residential Area, m^2	262	75202944.79	0	3140675.91	285942.76	364716.78

Min- Minimum, Obs.-Observations

Figure 3.15 shows the numbers of land use per Traffic Analysis Zone. Multiple zones had multiple land uses. This variable is an indicator of places that are more developed. Based on their land use, specific areas can attract certain traffic offenses and crimes. For instance, a zone with mixed land use, containing commercial, office, retail, and residential housing, attracts higher numbers of break and enter crimes compared to zones with open space recreational area and industry land use. The legend in Figure 3.15 represent the different land use types in a particular Traffic Analysis Zone. If a Traffic Analysis Zone has a commercial, recreational, and hospital land uses; it is assigned three (3), indicating the three different land use types indicated above.

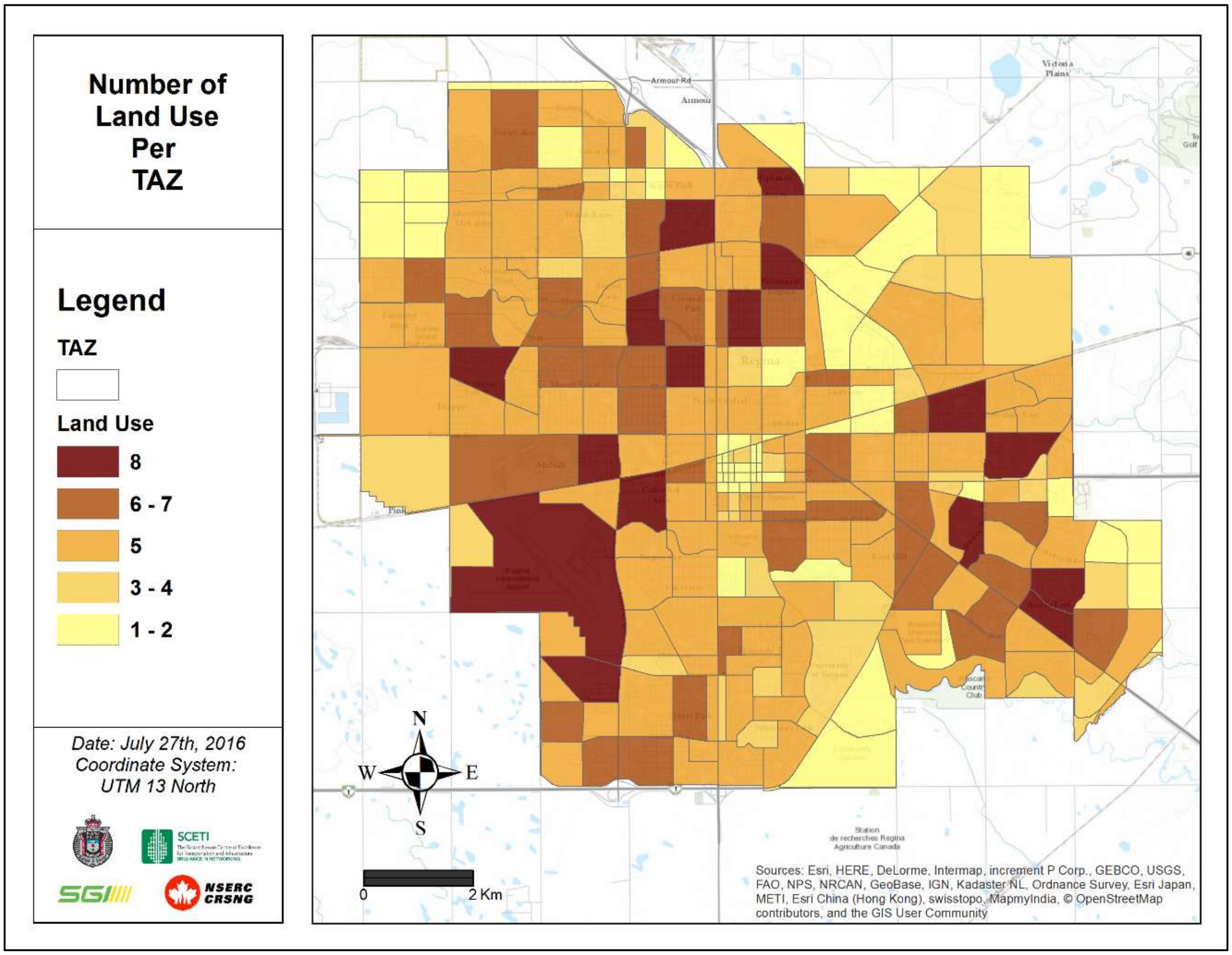


Figure 3.15: Map for Number of Land Use per Traffic Analysis Zone

3.4.4 Road Network and Infrastructure

Acquired road network data had numerous explanatory variables that would help predict collision as well as certain types of crimes that occur in close proximity to roads. Table 3.15 is a summary of statistics of road network variables aggregated on Traffic Analysis Zone level.

Table 3.15: Aggregated TAZ-Level Road Network and Infrastructure Data Statistics

Variable	Obs.	Total	Min.	Max.	Mean	Standard Deviation
Arterial Road Length (m)	262	145,654.95	0	2,618.06	555.93	565.05
Collector Road Length (m)	262	162,863.55	0	3,452.70	621.16	751.80
Expressway Length (m)	262	2,2037.00	0	2,029.00	84.11	262.64
Gravel Road Length (m)	262	35,762.00	0	4,879.00	136.50	488.31
Highway Length (m)	262	10,810.67	0	1,820.18	41.26	205.32
Local Road Length (m)	262	620,715.03	0	9,643.52	2,369.14	2,396.95
Private Road Length (m)	262	61,802.00	0	12,726.00	235.89	995.59
Ramp Length (m)	262	23,831.47	0	1,825.64	90.96	275.05
Right-Of-Way Length (m)	262	1,074.00	0	434.00	4.10	32.73
Total Road Segment Length (m)	262	1,084,721.83	87	16,354.95	4,140.16	3,299.99
Average Road Segment Length (m)	262	563,389.83	81.17	11,361.34	2,150.34	1,828.17
Roadway Length with Average Speed Limit (m)	262	22,277.33	40	536.17	85.03	55.84
Road Segment Length with posted Speed Limit 20km/hr	262	200.00	0	200.00	0.76	12.33
Road Segment Length with posted Speed Limit 30km/hr	262	1,179.00	0	523.00	4.50	45.01
Road Segment Length with posted Speed Limit 40km/hr	262	166,654.74	0	3,836.75	636.09	818.10
Road Segment Length with posted Speed Limit 50km/hr	262	824,639.38	0	14,405.84	3,147.48	2,703.98
Road Segment Length with posted Speed Limit 60km/hr	262	7,485.00	0	1,695.00	28.57	151.33

Table 3.15: Aggregated TAZ-Level Road Network & Infrastructure Data Statistics [cont'd]

Road Segment Length with posted Speed Limit 70km/hr	262	32,074.57	0	3,284.75	122.42	392.76
Road Segment Length with posted Speed Limit 80km/hr	262	34,125.00	0	3,677.00	130.25	431.64
Road Segment Length with posted Speed Limit 100km/hr	262	18,366.00	0	1,820.00	70.10	251.46
Number of three-leg intersections	262	2,362.00	0	66.00	9.02	10.02
Number of four-leg intersections	262	1,364.00	0	36.00	5.21	6.41
Number of five-leg intersections	262	4.00	0	1.00	0.02	0.12
Total Number of Intersections	262	3,725.00	0	79.00	14.22	13.63
Vehicle-Kilometer-Traveled (VKMT)	262	6,261,583.42	43.43	122,800.88	23,899.17	21,490.66
Annual Average Daily Traffic (AADT)	262	32,628,715.58	700	537,742.28	124,537.08	101,783.47
Total Lane Kilometer Traveled, TLKM (km)	262	1,084.72	0.09	16.35	4.14	3.30
Intersection Road Density, INTKD (Number of intersection per Total Lane Kilometers)	262	836.32	0	18.35	3.19	1.79
Proportion of three-leg intersections per Traffic Analysis Zone Area	262	14,248.66	0	100.00	54.38	34.22
Arterial Road Length Proportion (ALKP)	262	5,191.54	0	100.00	19.82	23.85
Local Road Length Proportion (LLKP)	262	12,413.35	0	100.00	47.38	30.64

Min- Minimum, Obs.-Observations, Max- Maximum

Intersections are conflict points that cause many collisions, and Figure 3.16 shows the distribution of the number of intersections per Traffic Analysis Zone. The legend represents the number of intersections in a Traffic Analysis Zone. Traffic Analysis Zones in both downtown areas and along Ring Road have high numbers of intersections. City of Regina's Ring Road is a high speed (100km/hr) highway that perimeters most of the city neighbourhoods, which implies that Traffic Analysis Zones that have the Ring Road as a boundary are high-risk for high occurrences of traffic collisions. Evidently, from the descriptive spatial analysis, most Fatal-Injury collisions occurred in Traffic Analysis Zones in downtown areas as well as zones along the Ring Road.

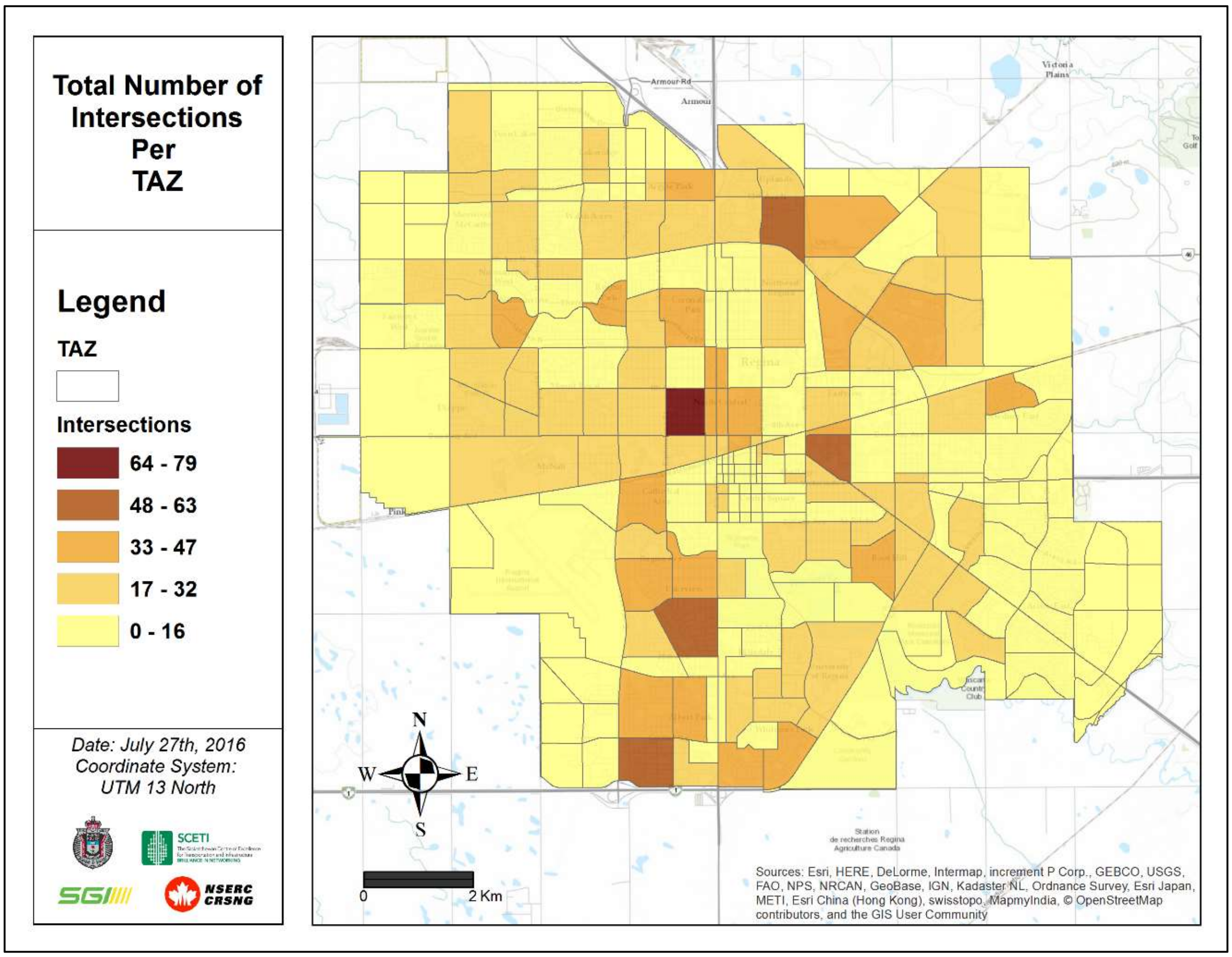


Figure 3.16: Map showing Aggregated Number of Intersections per Traffic Analysis Zone

Traffic exposure is strongly associated with collisions. As illustrated in Figure 3.17, which shows various Vehicle-Kilometer-Traveled levels, Traffic Analysis Zones in downtown areas and along Ring Road showed high numbers of Vehicle-Kilometer-Traveled. This finding supports the high occurrence of collisions. Furthermore, associated with high Vehicle-Kilometer-Traveled are high levels of crime in the downtown areas. Although, crime may not have a proven correlation with Vehicle-Kilometer-Traveled, the crime levels help inform the pattern at the modeling stage of this research. The University of Regina area showed the highest Vehicle-Kilometer-Traveled zone, which could be due to the high traffic volume of both private vehicles and public transit transporting students, faculty, and other staff.

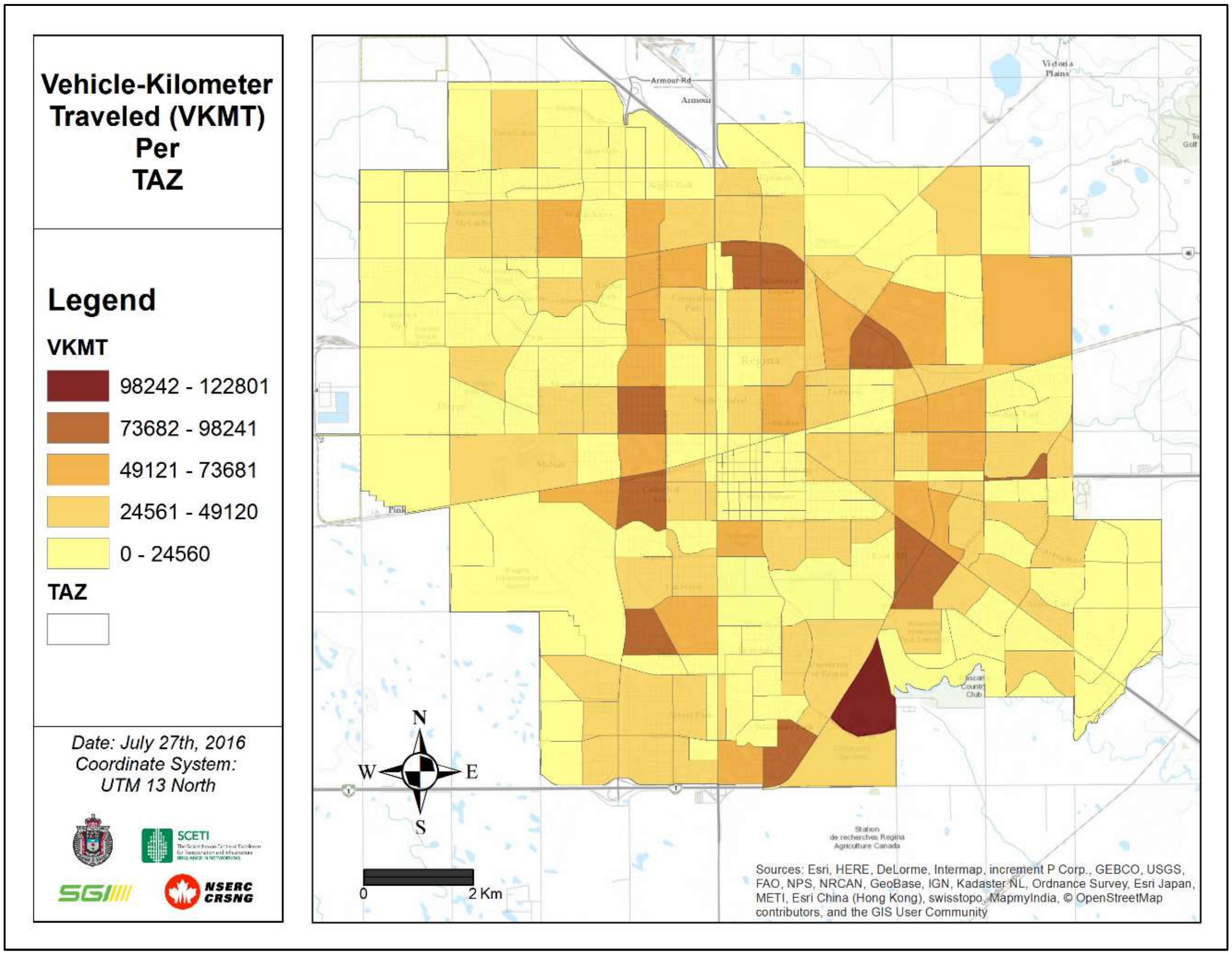


Figure 3.17: Spatial Distribution of Vehicle-Kilometer-Traveled Per Traffic Analysis Zone

3.4.5 Demographics

Statistics about the proportion of population age-groups in demographic data acquired are shown in Table 3.16. As can be seen in Table 3.16, the age-group, 25 to 44 years has the highest proportion of the City of Regina's population. This population age-group is the most active in any demographic. Figure 3.18 shows the spatial distribution of the active age group (25-44 years). A widespread distribution is evident across residential areas.

Table 3.16: Descriptive Statistics for TAZ-Level Demographic Data

Variable	Observations	Total	Min.	Maximum	Mean	Standard Deviation
Population aged 1 to 17	262	42,431	0	782	161	179.45
Population aged 18 to 24	262	21,145	0	339	80	84.12
Population aged 25 to 44	262	56,461	0	884	215	220.58
Population aged 45 to 64	262	53,726	0	911	204	214.47
Population aged 65 and above	262	27,443	0	763	104	125.84
Total Population	262	201,218	0	3,011	765	785.32
Number of graduate students	262	29,367	0	1,720	112	261.35

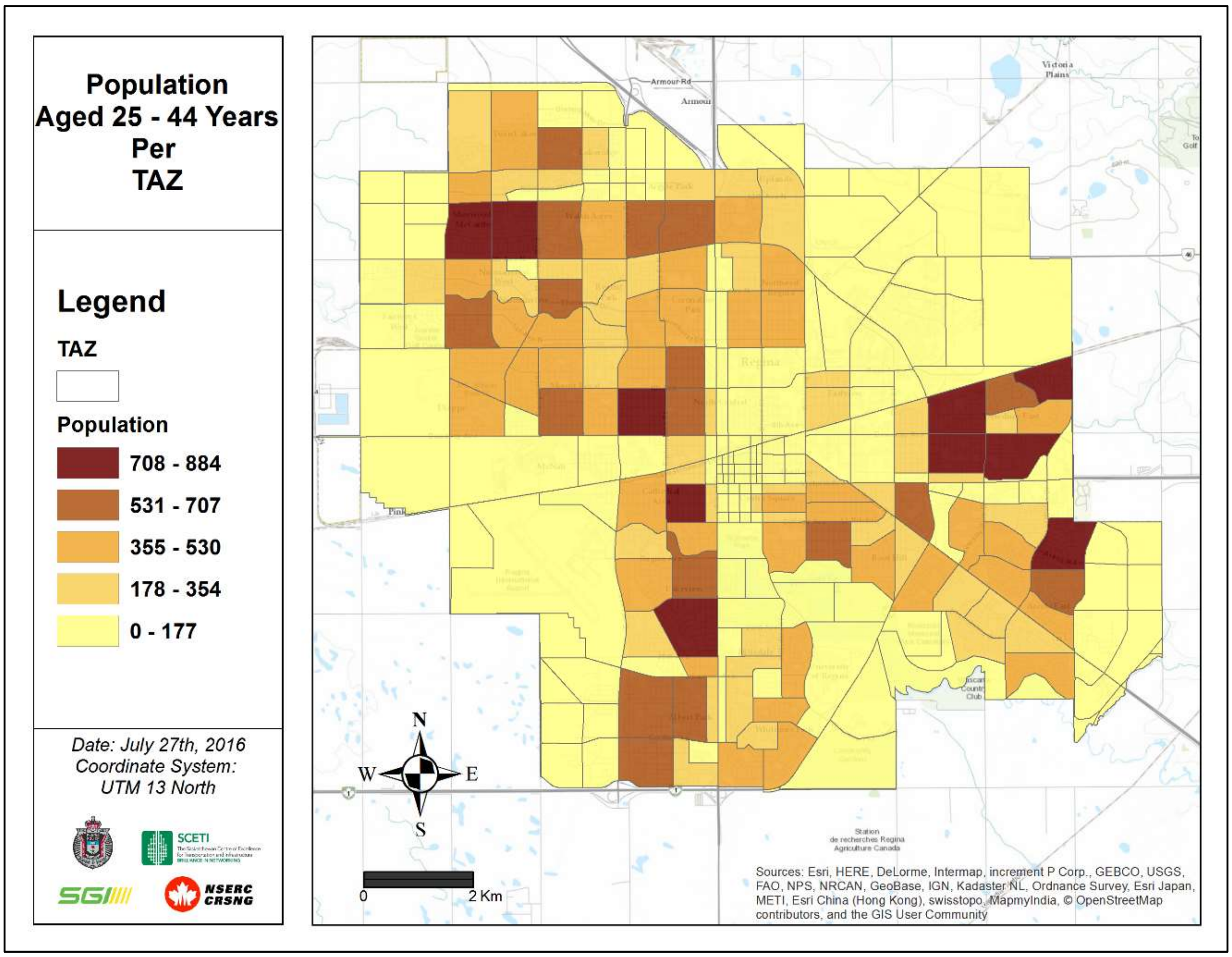


Figure 3.18: Spatial Distribution of Age Group 25-44 Years

The spatial distribution of population density per Traffic Analysis Zone is shown in Figure 24. The downtown area is the most densely populated area in the city. This is also an indicator of places that attract or generate trips, as well as some crime occurrence types.

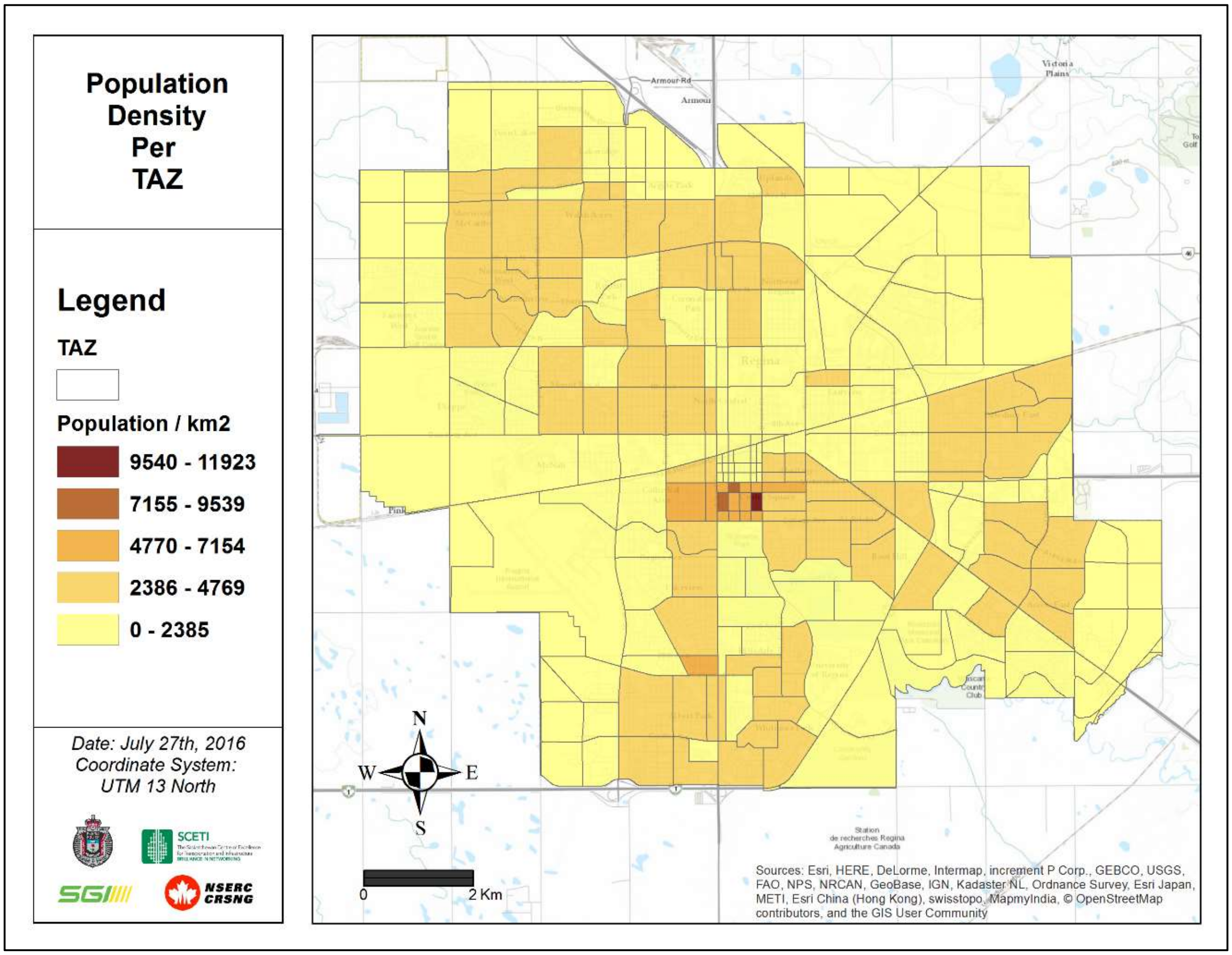


Figure 3.19: Spatial Distribution of Population Density by Traffic Analysis Zone

3.5 Chapter Summary

In this chapter, the City of Regina, which is the study area was discussed, and some statistics about the city were presented. The discussion examined various data sources and the ways in which these different databases were managed and aggregated. The issue of boundary data was also introduced. Through literature review, an equal assignment of boundary data to the number of Traffic Analysis Zones sharing the boundary was identified as providing the best predictive capability; thus, this option was employed in the research. Descriptive and some spatial aggregated level data statistics were presented for the different variable groups in this chapter. Other derived predictors from the acquired data were also presented.

CHAPTER 4 . RESEARCH METHODOLOGY

This chapter is dedicated to the methodology employed in this research. As outlined in Chapter One, one of the primary objectives of this research is to develop collision and crime prediction models for the City of Regina. Another objective is to develop a GIS based collision and crime mapping system to identify hotspots- places with high occurrence of collisions and or crimes. This research methodology is divided into three sections; pre-modeling, modeling and post-modeling, based on the various phases used in this research. The methodology adopted in this research is outlined in the flow chart shown in Figure 4.1 and as can be seen, the approach involves; data collection, data preparation, descriptive data analysis, model development, and hotspot identification.

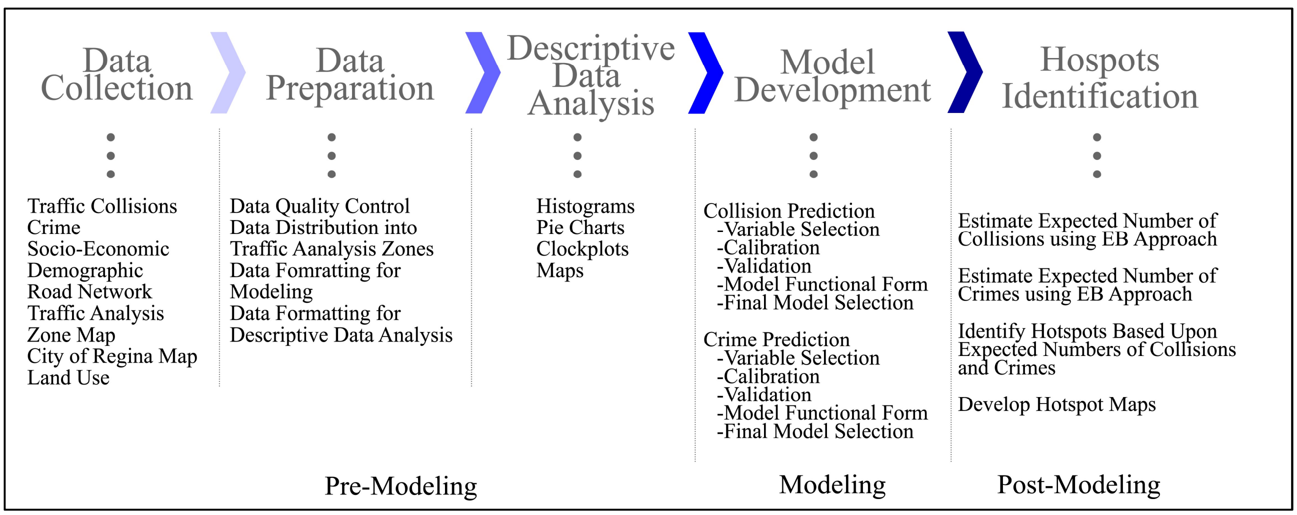


Figure 4.1: Methodology Flowchart

4.1 Pre-Modeling

This is the first part of the research and it involves the first three phases shown in the flowchart: data collection, data preparation, and descriptive data analysis. Data needed for this research were identified through literature review and previous research work in the area of macro-level prediction modeling. The dependent variables required for this research are traffic collisions and crimes. Independent variables include; traffic volume, socio-economic data, demographic data, road network characteristics, traffic analysis zones, land use data, and the City of Regina basemap. Data were then requested and acquired from various sources.

Various data quality controls were performed to ensure accuracy of data and subsequently accuracy of results of analysis from using such data. Data were then prepared into formats that can be readily used for modeling and descriptive analysis as well as spatio-temporal analysis. Models were developed to predict collisions and crimes per Traffic Analysis Zone. All variables both dependent and independent are assigned to various Traffic Analysis Zones based on their geographical location. For instance, collision and crimes are assigned to Traffic Analysis Zones in which they fall within. Traffic Analysis Zones are areas demarcated within a city with specific characteristics that are used for transportation and planning purposes by city engineers and planners. Data without location information such as longitude and latitude were excluded from modeling. Boundary data are variables that fall on boundaries of two or more Traffic Analysis Zones. Boundary data were assigned to neighbouring TAZs by equal proportion approach as previously explained in Chapter Three.

Descriptive data analysis was then performed for various variables per Traffic Analysis Zone to identify trends and distribution of variables across the City of Regina. Spatio-temporal exploratory analysis was also performed to determine the distribution of collision and crimes with respect to time.

Chapter Three of this thesis was dedicated to this stage of the research and results of the descriptive analysis was presented.

4.2 Modeling

At this step, a correlation matrix is performed to determine combination of variables to be kept in the same model. This helps to avoid having two highly correlated variables in the same model. The acquired data were randomly split into calibration and validation data. Calibration is a bigger percentage of the data and were used to model the dependent variables and determine estimates, standard errors as well as other model results. Validation data were used to authenticate the model.

Model form to be used was also identified through literature review. The model form used in this research is discussed in detail later in this Chapter. Several candidate models were then developed using the calibration data. Various goodness-of-fit tests were then applied to both calibration and validation data to test the predictive performance of the candidate models. If candidate models did not have good predictive performance, the whole process of modeling is repeated with different sets of variables and goodness-of-fit tests applied again. The best models were then selected based on their predictive performance. The various goodness-of-fit tests were explained in detail later in this Chapter.

4.3 Post Modeling

Using the estimates from the best selected models, the numbers of collisions and crimes were predicted per Traffic Analysis Zone. Empirical Bayes was then applied to the predicted numbers of collision and crime to estimate expected collision and crime numbers. Empirical Bayes was applied to compensate for the issue of regression-to-the-mean. Empirical Bayes approach is explained in detail later in this Chapter. The expected numbers of collisions and crimes were then used to create maps and hotspots (Traffic Analysis Zones) were identified. Expected numbers of collisions and crimes were then reported as final results of models.

4.4 Variable Selection

Correlation matrix was first developed to identify the most significant variables based on their correlation with the response variable (collision or crime). As expected, Vehicle-Kilometer-Traveled was the most significant explanatory variable. Variables with high correlations with the response variables were also noted as significant and candidate variables. Based on a forward selection procedure, model variables were selected. This approach begins with a model with the most significant variable (Vehicle-Kilometer-Traveled). The p-value of the Vehicle-Kilometer-Traveled is noted as being significant with a 5% significance level. The statistical significance of added variables was checked by their p-value.

It is first important to discuss an issue with regards to the null hypothesis significance testing, commonly referred to as p-value, before proceeding to results from models. There has been the issue of p-value being misleading in selecting variables in interpretation or prediction of events. Hauer (1991) studied a before and after case scenario on right-turn-on-red. In that study Hauer (1991) explored the definition of significance as suggested by statistics by comparing the recorded numbers of traffic collisions before and after the implementation of right-turn-on-red. Even though the difference in the expected and observed numbers of collisions was significant, right-turn-on-red was deemed insignificant based on the p-value. In a later study by Hauer (2004), three cases were studied, with right-turn-on-red case being one of them. The two other cases involved the safety effects of paving shoulders and increasing speed limits. Interestingly, all three studies showed similar results. Hauer (2004) concluded that in road safety cases, although a variable may not be statistically significant, the variable is significant in predicting or interpreting a dependent variable. Hauer (2004) added that the magnitude of estimates and their standard errors should be considered in choosing variables in a multi-variate regression modeling. In that regard, some variables may have p-values more than 0.05, which is the set significance level in this research; those variables may be kept in a model based on their improvement in the predictive capability of that model.

The overall predictive performance of the model was tested by a set of goodness-of-fit tests. These goodness-of-fit tests included the following: Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC), overdispersion parameter, Mean Squared Error (MSE), Freeman Turkey R-Squared (R2FT), Mean Prediction Bias (MPB), Mean Absolute Deviance (MAD), Mean Squared Prediction Error (MSPE), and Cumulative Residual (CURE) plots. These goodness-of-fit tests will be discussed later in this chapter and the results of the goodness-of-fit for the various candidate models are presented in Chapter Five.

In the case of collision modeling, upon adding any variable, both the calibration data (90% of total data) and validation data (10% of total data) were tested with the appropriate goodness-of-fit test to check the predictive performance of the model. If an added variable decreases the performance of the model, it is removed and the process was repeated for different sets of variables to obtain the best performing model. However, if two variables are highly correlated to the response variable and are also highly correlated to each other, these two explanatory variables are not kept in the same model to avoid the issue of multicollinearity. Another model with a different set of variables is then created to obtain the best fitting model.

The same approach was adopted in the creation of crime prediction models; 70% of total crime data were used for calibration and 30% of the total data were used for validation. The only difference in crime prediction was that no particular explanatory variable was known nor established in literature as the most significant in aggregated crime modeling. As such, models were created for different types of crimes, and the most significant variable for that crime type was determined before the forward selection is implemented. The ten goodness-of-fit tests listed earlier were used to check the predictive performance of the developed crime models.

4.5 Model Form

Various modeling techniques have been adopted by previous researchers in predicting collision and crimes. Some of these techniques are linear regression, ordinary least squares, Poisson regression, zero-inflated negative binomial, and negative binomial techniques. However, because collision and crime data are non-negative, the Ordinary Least Squares technique, which assumes a continuous variable cannot be used as a modeling technique. Most recent researchers have recommended Poisson and Negative Binomial. As has been established in literature review in Chapter Two, the Negative Binomial is the modeling technique adopted in this research.

The Negative Binomial model form used in this research has been already established in literature (Usman *et al*., 2011). The functional form is shown in Equation 4.1 below.

$$\mu_i = (\exp osure)^{\beta_1} * \exp\left(\beta_0 + \beta_2 x_{i1} + \beta_3 x_{i2} + \dots\dots\dots + \beta_k x_{ij}\right) \qquad 4.1$$

Where μ_i is the predicted number of collisions (response variable), x_{i1}, x_{i2}, x_{i3}, ……. x_{ij} are the predictors (explanatory variables), and $\beta_0, \beta_1, \beta_2, \beta_3, \dots\dots \beta_k$ are the parameters (coefficient) estimates from the regression model.

For easier computation and understanding, Equation 4.1 is converted into its linear form by taking log of both sides and re-arranging the variables:

$$\log(\mu_i) = \beta_0 + \beta_1 \log(\exp osure) + \beta_2 x_{i1} + \beta_3 x_{i2} + \dots\dots\dots + \beta_k x_{ij} \qquad 4.2$$

Having Vehicle-Kilometer-Traveled (VKMT) as the exposure variable, this equation can be rewritten as:

$$\log(\mu_i) = \beta_0 + \beta_1 \log(VKMT.) + \beta_2 x_{i1} + \beta_3 x_{i2} + \dots\dots\dots + \beta_k x_{ij} \qquad 4.3$$

where *VKMTi* is the VKMT for a particular zone *i*. The linear form was then used in modeling in R language and in Microsoft Excel computations. The model form shown in the above equations confirms suggestions by other researchers that the relationship between collisions and exposure, which is VKMT in this case, is not linear.

4.6 Empirical Bayes Approach

Empirical Bayes method compensates for the random fluctuations in collision incidents by using collision frequency and its predicted values based on the Collision Prediction Model. Road safety engineers have adopted this approach for road safety evaluations and to increase the accuracy of safety estimates (Hauer, 1992; Hauer, 1997; Hauer *et al*., 2002; Persaud *et al.*, 2002). As a result of site-specific safety improvements, the Empirical Bayes techniques can be used as evaluation tools for selected countermeasures. These techniques are based on weighted combinations of predicted collision by both the Collision Prediction Model and the observed number of collisions over the study period. The predicted collisions are a set of locations used as a reference and have a Collision Prediction Model similar to that of the target location. Empirical Bayes addresses regression-to-the-mean bias by the use of the before-and-after study with the Empirical Bayes method to adjust predicted collisions for regression-to-the-mean bias. As outlined by Shen (2007), the Empirical Bayes method is based on three assumptions:

1. Collision frequency at any location follows a Poisson distribution
2. The mean value for a set of systems can be estimated by a Gamma distribution
3. Yearly variations based on different factors are similar for all reference sites.

Various approaches have been suggested for estimating the expected collision frequency (Hauer, 1997; Mountain *et al*., 1992; Mountain and Fawaz, 1991; Wright *et al.*, 1988).

Multivariate regression methods can be used to calculate site-specific estimates of the mean collision frequency and variance of the set of reference locations (Hauer, 1992; Hauer, 1997). Equation 4.4 shows a mathematical representation of the Empirical Bayes technique (Hauer, 1997):

$$E\{k \mid K\} = \gamma * E\{k\} + (1-\gamma).K \qquad 4.4$$

where $E\{k/K\}$ is the estimate of the expected number of collisions at a site for the study period; γ is the weight factor and is greater than 0 but less than 1; $0 \leq \gamma \leq 1$; $E\{k\}$ is the predicted collision frequency at a reference site over the study period; K is the actual collision frequency at a site over the study period; and $E\{k\}$ refers to the predicted numbers from collision prediction models. The variance of the expected number of collisions is defined by (Hauer, 1992):

$$VAR\{k \mid K\} = (1-\gamma)E\{k \mid K\} \qquad 4.5$$

The weight factor therefore can be found by:

$$\gamma = \frac{1}{1+\frac{VAR\{k\}}{E\{k\}}} \quad 4.6$$

Because Empirical Bayes techniques heavily rely on the use of collision prediction models (CPM), CPMs must be carefully developed to be used with Empirical Bayes estimates.

4.7 Model Calibration

The total aggregated data were randomly split into two: calibration and validation data. Using 90% of the data to calibrate the model, this data were used to create an interaction with the response variable to determine coefficients for each variable. In determining the proportion of total data to be used in model calibration, different proportions were tested to check the performance of the calibrated model. These percentages (calibration-to-validation) were the following: 50-50, 60-40, 70-30, 75-25, 80-20, and 90-10. Predictive performance testing of different percentages of calibration-to-validation data showed that 90-10% gave the best fitting calibration, and, statistically, the data points (262) were small. Therefore, 90% was a reasonable proportion to use in model calibration to ensure the best fitting. In the case of crime prediction, a 70-30 (calibration-to-validation) proved much better predictive performance based on the goodness-of-fit tests results. The goodness-of-fit tests that are applied to the calibrated data, are presented in this section. These goodness-of-fit tests are cumulative residual plots, Akaike Information Criterion, Bayesian Information Criterion, overdispersion parameter, Mean Squared Error, and Freeman Turkey R-Squared (R2FT).

4.7.1 Cumulative Residual (CURE) Plots

Hauer and Bamfo (1997) proposed Cumulative Residual plots to compare different prediction models. Cumulative Residual plots are graphical representations of the predictive performance of a model across a range of data. Residuals are the differences between the observed and the predicted frequencies. In Cumulative Residual plots and in the case of collision modeling, the cumulative residuals are plotted against exposure, which is Vehivle-Kilometer-Traveled. Residuals above the zero line represent underestimations. On the other hand, residuals below the zero line represent a model that overestimates. Cumulative residuals that are within two standard deviations above and below the zero line at 95% confidence limit represent a good-fitting model.

Figures 4.2 and 4.3 show two different Cumulative Residual plots for two models, which were developed to predict Break and Enter crimes with population density on the x-axis and cumulative residual on the y-axis. In the Cumulative Residual plot, the black straight line is the zero line, representing the observed data. The green and red dotted lines represent two standard deviations above and below the zero line. The black zig-zag line represents the cumulative residual. Figure 4.26 shows a good fitting model for predicting Break and Enter crimes since the cumulative residual line is within two (negative and positive) standard deviations, and, generally, the line seems very close to the zero line. However, Figure 4.3 shows the Cumulative Residual plot for another model developed to predict Break and Enter crimes using the same data set. This figure represents a model that does not predict well and is not a good fit for Break and Enter crimes. Evidently, the x-axis scale illustrates that the magnitude for the Cumulative Residual plot in Figure 4.2 is much larger than the one in Figure 4.3.

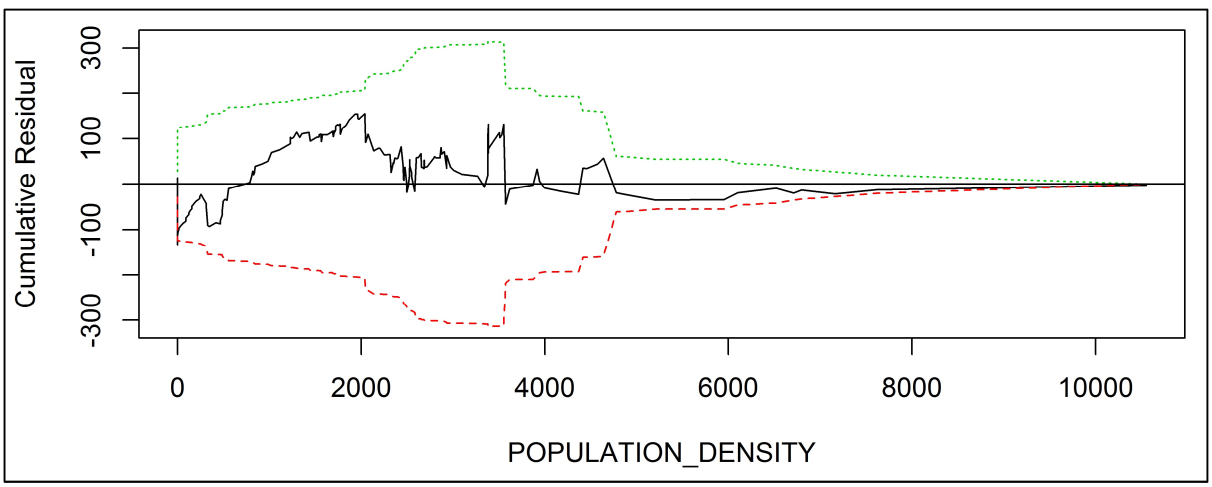


Figure 4.2: A Cumulative Residual Plot Showing a Good Fitting Model for Break and Enter Crimes

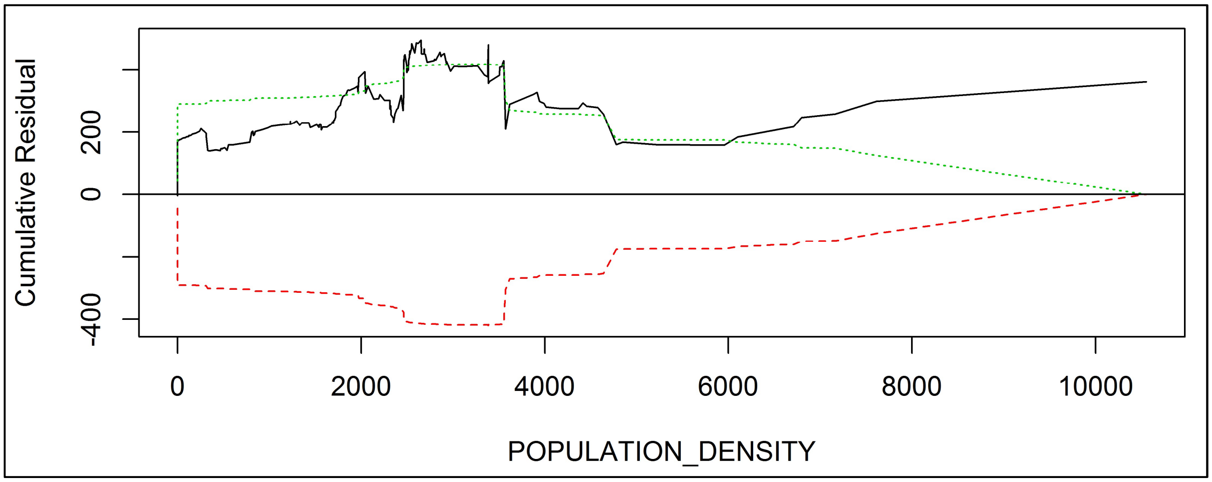


Figure 4.3: A Cumulative Residual Plot Showing a Poor Fitting Model for Break and Enter Crimes

4.7.2 Akaike Information Criterion (AIC)

Akaike Information Criterion measures a model's fit to the observed data. Lower values of Akaike Information Criterion indicate a better fitting model; however, Akaike Information Criterion does not indicate the model's performance. Akaike Information Criterion can be calculated using Equation 4.7.

$$AIC = 2\ x\ k - 2\ x \log(L) \qquad 4.7$$

where k is the number of variables in a model

L is the maximised value of the likelihood function for the estimated model (probability of the data given a model)

4.7.3 Bayesian Information Criterion (BIC)

Though comparable to Akaike Information Criterion, Bayesian Information Criterion takes into account the data sample size, which, in this case, is constituted of Traffic Analysis Zones. Bayesian Information Criterion can be estimated using Equation 4.8

$$BIC = k\ x \log(n) - 2\ x \log(L) \qquad 4.8$$

where L same meaning as in Akaike Information Criterion and n is the number of data sample size.

4.7.4 Overdispersion Parameter

Another measure of goodness-of-fit is the overdispersion parameter from a Negative Binomial regression model summary of results, indicating estimates and other statistics. Overdispersion parameter measures the heterogeneity or variance in the dataset. Larger values of the overdispersion parameter indicate an increase in variance of the dataset. Thus, smaller values of overdispersion parameter are typically preferred (Washington *et al.*, 2005). In this research, the statistical package used in developing the model was *R*. The output from *R* provides the inverse of the overdispersion parameter: the dispersion parameter. Therefore, in this case, larger values of dispersion parameter are preferred.

4.7.5 Mean Squared Error (MSE) and Root Mean Squared Error (RMSE)

Suggested by Washington *et al.* (2005), Mean Squared Error (MSE) is a statistical test that is applied to the calibration data and measures the mean or average of the squares of errors associated with a predictor. The result of this test is essentially the square of the difference between the predicted and observed values. The equation below illustrates Mean Squared Error.

$$MSE = \frac{\sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)^2}{n-p} \qquad 4.9$$

where $\widehat{Y_i}$ is the predicted number of events in Traffic Analysis Zone i, Y_i is the observed number of events, and n is the sample size of data, and p is the number of explanatory variables in the model.

Smaller values (values closest to zero) of Mean Squared Error indicate a better predictive performance of a model. The square root of Mean Squared Error is the Root Mean Squared Error. Root Mean Squared Error is the standard deviation of the prediction errors, measuring the data spread around the regression. A Root Mean Squared Error value that is closer to zero implies a model that predicts with less errors.

4.8 Model Validation

Validation data were the remaining 10% in the case of collisions and 30% in the case of crimes. The reason behind the data split is to validate the model created. By using the estimates or coefficients from models created, event (collisions and crimes) frequencies were predicted using the validation data. These frequencies were then compared to the observed frequencies of events (collisions and crimes). As opposed to an explanatory model where 100% data can be used in model calibration, it is important to check the data in predictive modeling because the validation data had no interaction in the coefficient estimation. The various goodness-of-fit tests are described below. Furthermore, because no individual test can ascertain the accuracy of a model, multiple tests are used to validate the model, thus ensuring the selection of the best model. The descriptions below refer to both collision and crime.

4.8.1 Mean Square Prediction Error (MSPE)

Mean Squared Prediction Error tests the errors associated with the predicted event frequencies. However, this test is performed on the validation data. Comparable to Mean Squared Error, values closest to zero are preferable. The equation below illustrates Mean Squared Prediction Error (Hamidi *et al*. 2010).

$$MSPE = \frac{\sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)^2}{n} \qquad 4.10$$

where $\widehat{Y_i}$ is the predicted event frequency in Traffic Analysis Zone i, Y_i is the observed event frequency, and *n* is the sample size of data.

Further important information that can be derived using Mean Squared Error and Mean Squared Prediction Error is under-fitting or over-fitting of the model. A model with MSPE>MSE is overfitting. On the other hand, a model with MSPE<MSE is an under-fitting model. This implies that the closer Mean Squared Prediction Error and Mean Squared Error values are for a model, the better the predictive capability of that model.

4.8.2 Freeman Turkey R-Squared (R2FT)

The Freeman Turkey R-Squared (R2FT) is applied to both the calibration and validation data, as the values from both datasets are compared. Larger values of Freeman Turkey R-Squared are preferable. Equations 4.11 to 4.14 illustrate how to estimate Freeman Turkey R-Squared (Hamidi *et al.*, 2010; Washington *et al*., 2005).

$$R_{FT}^{2} = \frac{\sum_{i=1}^{n}\left(f_i - \hat{f}_i\right)^2 - \sum_{i=1}^{n}\hat{e}_i^2}{\sum_{i=1}^{n}\left(f_i - \hat{f}_i\right)^2} \qquad 4.11$$

where

$$f_i = \sqrt{Y_i} + \sqrt{Y_i + 1} \qquad 4.12$$

$$\hat{f} = \frac{\sum_{i=1}^{n}\left(\sqrt{Y_i} + \sqrt{Y_i + 1}\right)}{n} \qquad 4.13$$

$$\hat{e} = \sqrt{Y_i} + \sqrt{Y_i + 1} - \sqrt{4\hat{Y}_i + 1} \qquad 4.14$$

$Y_i, n, and\ \widehat{Y_\iota}$ have the same meanings as previously explained.

4.8.3 Mean Prediction Bias (MPB)

Applied to the validation data, Mean Prediction Bias measures the magnitude and direction of the average model bias. Smaller values of Mean Prediction Bias indicate a better fit of the predicted numbers of events to the observed numbers of events. The value is derived by the summation of differences between observed and predicted event frequencies divided by sample size (Traffic Analysis Zones in this case). Mean Prediction Bias is calculated with Equation 4.15 (Hamidi *et al.*, 2010; Washington *et al.*, 2005).

$$MPB = \frac{\sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)}{n} \qquad 4.15$$

The symbols in the equation denote same explanation given earlier.

Mean Prediction Bias can be either negative, in the case of underestimation, or positive, in the case of overestimation. The smaller the absolute value of Mean Prediction Bias is, the better the predictive capability of the model.

4.8.4 Mean Absolute Deviation (MAD)

Mean Absolute Deviation is applied to the validation data, measuring the magnitude of the average model bias. Unlike MPB, the value of Mean Absolute Deviation can only be positive. The model with the smallest Mean Absolute Deviation value for a dataset is considered the best fit. Mean Absolute Deviation is estimated using Equation 4.16 (Hamidi *et al.*, 2010; Washington *et al.*, 2005).

$$MAD = \frac{\sum_{i=1}^{n} \left| \hat{Y}_i - Y_i \right|}{n} \tag{4.16}$$

4.9 Chapter Summary

The concept of collision prediction models was explained as provided in the Highway Safety Manual and literature in this chapter. The issue of regression-to-mean bias was examined, and the empirical Bayes approach was discussed as the solution to prevent the effects of regression-to-the-mean bias. In addition to the mathematical presentation and the underlying principles of Empirical Bayes method, the model form for this research, background information about Negative Binomial technique, and model calibration and validation were presented in this chapter. In the case of collisions, 90% of total data were used to calibrate, and 70% of total data were used in crime model calibration. Therefore, 10% of total collision data were used to validate the developed models, and, in the same manner, 30% of the remaining total crime data were used to validate the created crime models.

The various goodness-of-fit tests were discussed. These goodness-of-fit tests were grouped into two: those applied to the calibration data, and those applied to the validation data. goodness-of-fit tests applied to the calibrated data included Cumulative Residual (CURE) plots, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), overdispersion parameter, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the Freeman Turkey R-Squared (R2FT). Mean Squared Prediction Error (MSPE), Freeman Turkey R-Squared (R2FT), Mean Prediction Bias)MPB), and Mean Absolute Deviation (MAD) are among the goodness-of-fit tests applied to the validation data.

CHAPTER 5 . RESULTS AND DISCUSSION

Results from models created are presented here. For each response variable, a number of different models were created and various goodness-of-fit tests were applied to determine the best fitting model. Maps created from these models will be presented. It is worthy of mention that final maps presented in this chapter represent estimates from the Empirical Bayes technique. Hotspot maps were then created for observed, predicted, and expected frequencies of collisions. Hotspot maps representing observed collision frequencies are presented in appendix C and hotspot maps for predicted collision frequencies are presented in appendix E of this thesis.

5.1 Collision Prediction Model Result

In total, there were 76 candidate variables considered in creating collision prediction models. For each severity type, several candidate models were created. Those models were shortlisted to 10 candidate models, based on the results from the goodness-of-fit test applied to them. These 10 models were further compared, and, using the goodness-of-fit tests as basis, the best model was selected. Since multiple goodness-of-fit tests are being used, some models may exhibit a good fit based on one goodness-of-fit, but others would show consistency in terms of multiple goodness-of-fit tests; thus, such models will be considered over others.

5.1.1 Total Collision

Over 200 Negative Binomial regression models were developed to predict Total collisions. Goodness-of-fit tests were then applied to these models. Based on the results of these goodness-of-fit tests, the top 10 models with the best predictive capability were selected as the top candidate models. The results of the Negative Binomial regression for these top ten models are presented in Appendix H of this thesis. Results of the various goodness-of-fit test applied to both the calibration and validation data for the top ten total collisions models are presented in Table 5.1. Based on the various goodness-of-fit tests and their comparisons, model 1 (highlighted in Table 5.1) was selected as the best fitting model. Table 5.2 provides the different sets of covariates for the top 10 Total collision models and their corresponding p-values.

Table 5.1: Summary of Results of goodness-of-fit Tests for Total Collisions

Model	AIC	BIC	Dispersion Parameter	MSE	MSPE	R2FT		MPB	MAD	RMSE
						Calibration Data	Validation Data			
1	**2469.08**	**2500.26**	**1.83**	**3280.90**	**3071.08**	**0.650**	**0.729**	**-15.301**	**39.18**	**57.28**
2	2467.50	2498.68	1.82	4641.12	6260.21	0.610	0.507	-15.966	57.34	68.13
3	2528.85	2563.49	1.43	4563.65	5335.18	0.606	0.546	-5.310	52.08	67.55
4	2528.52	2552.77	1.39	3686.03	3843.40	0.598	0.614	-8.988	45.75	60.71
5	2530.19	2554.44	1.38	4039.73	4427.61	0.581	0.579	-6.514	49.46	63.56
6	2529.01	2553.25	1.39	3905.70	4644.76	0.592	0.568	-7.348	51.84	62.50
7	2557.50	2585.21	1.23	4458.03	4668.03	0.533	0.494	-8.943	52.22	66.77
8	2513.02	2544.20	1.52	3231.40	4805.23	0.633	0.568	-5.078	53.33	56.85
9	2476.38	2507.55	1.77	3945.79	4500.90	0.622	0.627	-13.919	48.37	62.82
10	2474.58	2502.29	1.77	3844.01	4556.00	0.625	0.625	-13.908	48.60	62.00

AIC- Akaike Information Criterion,

BIC- Bayesian Information Criterion,

MSE- Mean Squared Error,

MSPE-Mean Squared Prediction Error,

R2FT- Freeman Turkey R-Squared,

MPB- Mean Prediction Bias,

MAD- Mean Absolute Deviation,

RMSE- Root Mean Squared Error

Dispersion parameter was obtained from calibration data.

Table 5.2: Top 10 Total Collision Models Covariates and their P-Values

Covariate	P-value
Model 1	
Intercept	0.0014
logVKMT	< 0.001
I3WP	0.0561
INTKD	< 0.001
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	< 0.001
FOUR_LEG_INTERSECTIONS	0.0011
SEGMENT_80KMHR	0.0413
LOW_DENSITY_RESIDENTIAL_AREA_PROP	0.0116
Model 2	
Intercept	< 0.001
logVKMT	< 0.001
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	< 0.001
POPULATION_DENSITY	< 0.001
FOUR_LEG_INTERSECTIONS	< 0.001
THREE_LEG_INTERSECTIONS	0.4416
POPULATION_18TO24	0.1285
POPULATION_45TO64	0.0120
Model 3	
Intercept	< 0.001
logVKMT	< 0.001
FOUR_LEG_INTERSECTIONS	< 0.001
INTERSECTION_DENSITY	< 0.001
SEGMENT_80KMHR	< 0.001
THREE_LEG_INTERSECTIONS	0.1481
LOW_DENSITY_RESIDENTIAL_AREA_PROP	0.1131
POPULATION_18TO24	0.0471
POPULATION_45TO64	0.0694
Model 4	
Intercept	< 0.001
logVKMT	< 0.001
FOUR_LEG_INTERSECTIONS	< 0.001
INTERSECTION_DENSITY	< 0.001
SEGMENT_80KMHR	< 0.001
LOW_DENSITY_RESIDENTIAL_AREA_PROP	0.0147
Model 5	
Intercept	< 0.001
logVKMT	< 0.001
INTERSECTION_DENSITY	< 0.001
SEGMENT_80KMHR	< 0.001
POPULATION_45TO64	< 0.001
FOUR_LEG_INTERSECTIONS	< 0.001

Table 5.2 Top 10 Total Collision Models Covariates and their P-values [cont'd]

Covariate	P-Value
Model 6	
Intercept	< 0.001
logVKMT	< 0.001
SEGMENT_80KMHR	< 0.001
FOUR_LEG_INTERSECTIONS	< 0.001
INTERSECTION_DENSITY	0.0013
THREE_LEG_INTERSECTIONS	0.0115
Model 7	
Intercept	0.0019
logVKMT	< 0.001
THREE_LEG_INTERSECTIONS	0.2475
POPULATION_45TO64	0.9684
LOW_DENSITY_RESIDENTIAL_AREA_PROP	0.2185
FOUR_LEG_INTERSECTIONS	< 0.001
INTERSECTION_DENSITY	< 0.001
Model 8	
Intercept	< 0.001
logVKMT	< 0.001
I3WP	0.0406
ALKP	0.0150
INTKD	< 0.001
FOUR_LEG_INTERSECTIONS	0.0896
SEGMENT_80KMHR	< 0.001
THREE_LEG_INTERSECTIONS	0.0052
Model 9	
Intercept	0.0022
logVKMT	< 0.001
SEGMENT_80KMHR	0.0529
FOUR_LEG_INTERSECTIONS	< 0.001
INTERSECTION_DENSITY	0.0249
THREE_LEG_INTERSECTIONS	0.0061
ALKP	0.6301
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	< 0.001
Model 10	
Intercept	0.0015
logVKMT	< 0.001
SEGMENT_80KMHR	0.0430
FOUR_LEG_INTERSECTIONS	< 0.001
INTERSECTION_DENSITY	0.0233
THREE_LEG_INTERSECTIONS	0.0022
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	< 0.001

As can be seen from Table 5.2, all 10 models had log-transformed Vehicle-Kilometer-Traveled (logVKMT) as one of the covariates. Being the main exposure variable, Log-transformed Vehicle-

Kilometer-Traveled was the most important variable in predicting collisions. 4-leg intersections was also present in all 10 candidate models as this was one of the most significant predictors of Total collisions. In addition to log-transformed Vehicle-Kilometer-Traveled and 4-leg intersections, the different models had different sets of predicting covariates. Most of the covariates in the different models were significant with p-values below 0.05. Variables with p-values greater than 0.05 that were kept in their respective models significantly improved the predictive capability of models.

Table 5.3 shows the Negative Binomial regression model results for the selected Total collision model. All variables in the selected Total collision model were significant with p-values less than 0.05 except for three-leg intersection density, which has a value of 0.056; even though it is slightly above the significant value of 0.05, it improved the predictive performance of the model significantly and, as such, was maintained in the model. All the variables were not highly correlated with each other, and that avoided the issue of collinearity. Figure 5.1 shows the Cumulative Residual plot for the selected Total collision model, and that showed a good prediction over almost the entire range of Vehicle-Kilometer-Traveled. Cumulative residuals were mostly within the +2 and -2 standard deviations, which illustrated the best fitting model for total collisions. As can be seen from Table 5.3, log-transformed Vehicle-Kilometer-Traveled (logVKMT), intersection road density (INTKD), and 4-leg (4-way) intersections are positively associated with total collisions.

Increase in log-transformed Vehicle-Kilometer-Traveled indicates an increase in total collisions, and this can be explained by the fact that the higher the traffic volume, the more chances there are for collisions to occur. Places with lower traffic volumes or exposure are expected to experience fewer numbers of collisions. At intersections, road users make decisions to either go straight or turn; these decisions lead to conflict points at intersections, and these conflict points then result in collisions. Therefore, the more conflict points per road length, the more collisions are expected. This finding explains why an increase in INTKD indicates an increase in total collisions. Similarly, 4-leg intersections are conflict points in a road network and influence collisions. Therefore, an increase in the number of 4-leg intersections would indicate an increase in total collisions.

On the other hand, increases in the proportion of 3-leg intersections (I3WP), proportion of urban holding residential areas, road segments with posted speed limit of 80 km/hr, and proportion of low density residential areas result in a decrease in total collisions. Even though intersections are conflict points, the number of conflict points increases with a higher number of legs of the intersection. Three-leg intersections have fewer conflict points, and, therefore, an increase in I3WP has a decreasing effect on the number of total collisions. Urban holding residential areas are lands that have plans for future development, and, as such, there no road corridors. Therefore, increase in the proportion of urban holding residential areas has a decreasing effect on total collisions. Road segments with posted speed limits of 80 km/hr are high speed roadways with multiple lanes and minimal intersections, implying fewer conflict points and enough shoulder lanes. These road segments, therefore, have a reducing effect on total collisions. Low density residences have fewer people and less traffic, which results in a reduction in total collisions.

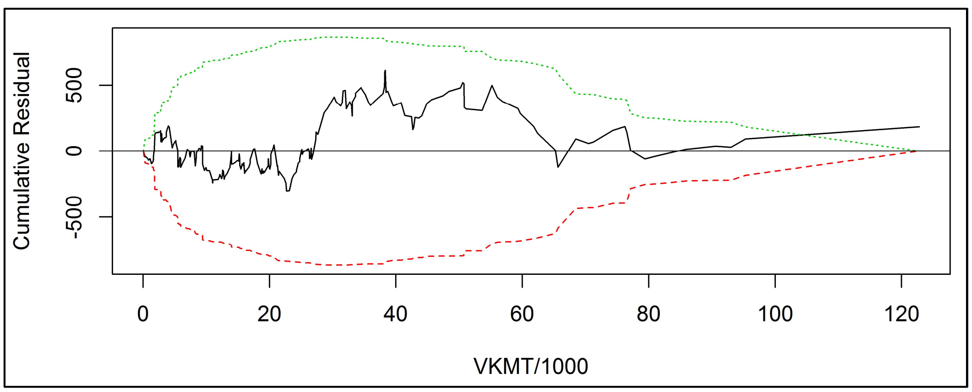


Figure 5.1: Cumulative Residual Plot for Model 8: Total Collisions

Table 5.3: Total Collisions Negative Binomial Regression Model Results

Covariate	Estimate	Standard Error	Z value	Pr(>|z|)	Model Statistics
Intercept	-1.73E+00	5.39E-01	-3.20219	0.0014	Dispersion Parameter = 1.827
logVKMT	4.58E-01	5.93E-02	7.70868	< 0.001	Standard Error = 0.186
I3WP	-3.58E-03	1.87E-03	-1.91043	0.0562	Log-likelihood = -2451.082
INTKD	1.21E-01	3.32E-02	3.63535	< 0.001	
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	-2.31E+00	2.80E-01	-8.26933	< 0.001	
FOUR_LEG_INTERSECTIONS	3.61E-02	1.11E-02	3.25450	0.0011	
SEGMENT_80KMHR	-2.94E-04	1.44E-04	-2.04095	0.0413	
LOW_DENSITY_RESIDENTIAL_AREA_PROP	-4.74E-01	1.88E-01	-2.52487	0.05116	

Equation 5.1 is a mathematical representation of the selected Total collision model. Equations representing the mathematical form of the top 10 total collision models are presented in Appendix H.

$$TOT = N \times \exp(-1.73) \times \exp\begin{pmatrix} (\log VKMT \times 0.458) + (I3WP \times -3.58 \times 10^{-3}) + (INTKD \times 0.121) + \\ (URBAN\_HOLDING\_RESIDENTIAL\_AREA\_PROP \times -2.31) + \\ (FOUR\_LEG\_INTERSECTIONS \times 0.0361) + (SEGMENT\_80KMHR \times -2.94 \times 10^{-4}) + \\ (LOW\_DENSITY\_RESIDENTIAL\_AREA \times -0.47) \end{pmatrix} \qquad 5.1$$

N is the number of years.

Using the estimates from the selected Total collision model, the predicted numbers of total collisions were calculated for all Traffic Analysis Zones. Empirical Bayes method was then employed to estimate the expected number of Total collisions per Traffic Analysis Zone. Figure 5.2 is a map showing the expected number of total collisions per Traffic Analysis Zone.

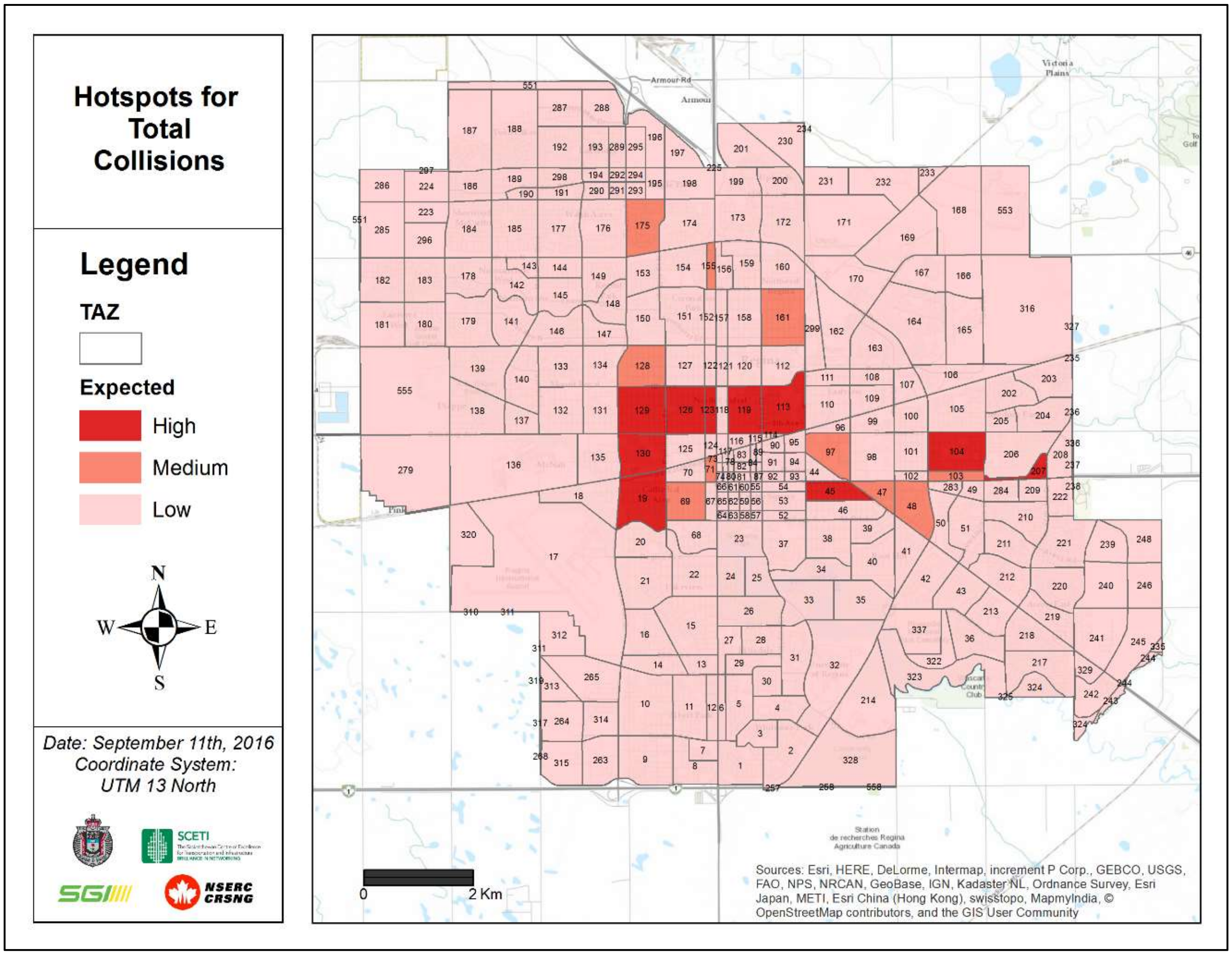


Figure 5.2: Hotspot Map for Expected Numbers of Total Number of Collisions

Traffic Analysis Zones were then ranked based on the number of expected Total collisions. Traffic Analysis Zones with the 10 highest collision frequencies were then grouped as the high hotspots; Traffic Analysis Zones with collision frequencies between ranked numbers 11 and 20 were grouped as medium hotspots; and the remaining Traffic Analysis Zones were grouped as low hotspots. This approach was adopted in determining hotspots for all maps. Table 5.4 gives the range of values grouped as high, medium or low as labelled in the map legend for Total collisions.

Table 5.4: Expected Numbers of Total Collisions per Traffic Analysis Zone Legend

Map Legend	Number of collisions	Number of Traffic Analysis Zones
High	283.34 – 424.27	10
Medium	239.52 – 273.36	10
Low	Less than 239	242

The list of the top 10 Total collision TAZ hotspots are also presented in Table 5.5. Table 5.5 provides the numbers of expected total collisions for each of the top 10 TAZ hotspots.

Table 5.5: Top 10 Total Collision Hotspots

Rank	Traffic Analysis Zone Number	Number of Expected Total Collisions
1	126	424.27
2	113	404.48
3	129	403.32
4	123	366.36
5	19	362.14
6	119	351.81
7	130	345.91
8	104	311.24
9	45	284.81
10	207	283.38

5.1.2 Fatal-Injury Collision

Out of the several developed models to predict Fatal-Injury collisions, the best 10 were selected based on their predictive performance or fit to the data. Table 5.6 presents the results from the various goodness-of-fit tests for the top 10 Fatal-Injury collisions employed in this research. Model 1 (highlighted in Table 5.6) showed the best fit to Fatal-Injury collisions. Model 1 was therefore, the selected Fatal-Injury collisions model. Table 5.7 provides the predictor variables in the top 10 Fatal-Injury collisions and their p-values. Log-transformed Vehicle-kilometer-Traveled and 4-leg intersections were the most significant predictors and were present in all 10 models. Most predictor variables had p-values less than 0.05. In few instances, variables with p-values greater than 0.05 were kept in their respective models because they significantly improved the predictive performance of models.

Table 5.6: Summary of Result of goodness-of-fit Tests for Fatal-Injury Collisions

Model	AIC	BIC	Dispersion Parameter	MSE	MSPE	R2FT		MPB	MAD	RMSE
						Calibration Data	Validation Data			
1	**1753.24**	**1787.88**	**2.15**	**239.36**	**182.83**	**0.60**	**0.70**	**-4.55**	**8.86**	**15.47**
2	1760.93	1788.64	2.03	230.78	191.96	0.59	0.68	-3.37	9.47	15.19
3	1806.79	1831.04	1.67	228.82	204.70	0.56	0.64	-1.93	10.44	15.13
4	1806.59	1834.30	1.68	423.25	207.57	0.43	0.62	-1.98	10.69	20.57
5	1805.67	1829.92	1.67	238.56	203.76	0.55	0.63	-1.80	10.57	15.45
6	1803.28	1830.99	1.72	235.43	180.65	0.57	0.68	-2.41	9.28	15.34
7	1830.10	1861.28	1.51	258.10	198.64	0.52	0.59	-2.85	9.78	16.07
8	1754.41	1785.59	2.11	232.05	215.33	0.60	0.62	-3.66	10.46	15.23
9	1760.14	1791.32	2.06	224.00	136.37	0.60	0.77	-3.47	7.29	14.97
10	1754.16	1785.34	2.12	245.20	173.33	0.59	0.72	-3.77	8.74	15.66

Table 5.7: Top 10 Fatal-Injury Collision Models Covariates and their P-Values

Covariate	P-Value
Model 1	
Intercept	< 0.001
logVKMT	< 0.001
SEGMENT_80KMHR	0.0157
SEGMENT_70KMHR	0.0840
FOUR_LEG_INTERSECTIONS	< 0.001
INTERSECTION_DENSITY	0.0017
THREE_LEG_INTERSECTIONS	0.0317
log1p(COLLECTOR_LENGTH)	0.0039
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	< 0.001
Model 2	
Intercept	< 0.001
logVKMT	< 0.001
SEGMENT_80KMHR	0.0027
FOUR_LEG_INTERSECTIONS	0.0061
INTERSECTION_DENSITY	0.0037
THREE_LEG_INTERSECTIONS	< 0.001
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	< 0.001
Model 3	
Intercept	< 0.001
logVKMT	< 0.001
SEGMENT_80KMHR	< 0.001
FOUR_LEG_INTERSECTIONS	0.0071
INTERSECTION_DENSITY	< 0.001
THREE_LEG_INTERSECTIONS	< 0.001
Model 4	
Intercept	< 0.001
logVKMT	< 0.001
FOUR_LEG_INTERSECTIONS	< 0.001
INTERSECTION_DENSITY	< 0.001
SEGMENT_80KMHR	< 0.001
SEGMENT_60KMHR	< 0.001
POPULATION_45TO64	< 0.001
Model 5	
Intercept	< 0.001
logVKMT	< 0.001
INTERSECTION_DENSITY	< 0.001
SEGMENT_80KMHR	< 0.001
POPULATION_45TO64	< 0.001
FOUR_LEG_INTERSECTIONS	< 0.001

Table 5.7: Top 10 Fatal-Injury Collision Models Covariates and their P-Values [cont'd]

Covariate	P-Value
Model 6	
Intercept	< 0.001
logVKMT	< 0.001
SEGMENT_80KMHR	< 0.001
FOUR_LEG_INTERSECTIONS	0.0018
INTERSECTION_DENSITY	< 0.001
THREE_LEG_INTERSECTIONS	0.0678
log1p(COLLECTOR_LENGTH)	0.0156
Model 7	
Intercept	< 0.001
logVKMT	< 0.001
THREE_LEG_INTERSECTIONS	0.5544
POPULATION_45TO64	0.9285
LOW_DENSITY_RESIDENTIAL_AREA_PROP	0.1141
FOUR_LEG_INTERSECTIONS	< 0.001
COLLECTOR_LENGTH	0.0275
INTERSECTION_DENSITY	< 0.001
Model 8	
(Intercept)	< 0.001
logVKMT	< 0.001
FOUR_LEG_INTERSECTIONS	< 0.001
INTERSECTION_DENSITY	0.01360
SEGMENT_80KMHR	< 0.001
SEGMENT_60KMHR	< 0.001
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	< 0.001
POPULATION_45TO64	< 0.001
Model 9	
Intercept	< 0.001
logVKMT	< 0.001
I3WP	0.0116
INTKD	0.0017
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	< 0.001
FOUR_LEG_INTERSECTIONS	0.0240
SEGMENT_80KMHR	0.0024
LOW_DENSITY_RESIDENTIAL_AREA_PROP	0.0034
Model 10	
Intercept	< 0.001
logVKMT	< 0.001
SEGMENT_80KMHR	0.0056
FOUR_LEG_INTERSECTIONS	< 0.001
INTERSECTION_DENSITY	0.0060
THREE_LEG_INTERSECTIONS	0.0510
log1p(COLLECTOR_LENGTH)	0.0023
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	< 0.001

There were 8 variables in the selected Fatal-Injury collision model, and Table 5.8 is a summary of Negative Binomial regression model estimates for the selected Fatal-Injury model. Negative Binomial regression models results for the top 10 Fatal-Injury collision models are presented in Appendix H. All 8 variables were neither highly correlated with each other nor with log-transformed Vehicle-Kilometer-Traveled. All the variables, except road segment with posted speed limit of 70 km/hr, had p-values less than 0.05, making seven of them significant. Collision Prediction Models for Fatal-Injury collisions predicted quite better than that for total collisions. +2 and -2 standard deviation values were smaller and Figure 5.3 is the Cumulative Residual plot for the selected model. The Cumulative Residual plot showed a good prediction within the +2 and -2 standard deviations over the entire Vehicle-Kilometer-Traveled range. Interestingly, the cumulative residual line almost converges with the zero line, meaning the cumulative predicted numbers of Fatal-Injury collisions were almost the same as the observed number of Fatal-Injury collisions. The Cumulative Residual plot for all top 10 Fatal-Injury collision models are presented in the Appendix D. Negative Binomial regression model results for these top 10 Fatal-Injury collisions are also presented in Appendix H.

Log-transformed Vehicle-Kilometer-Traveled (logVKMT), road segments with posted speed limit of 70 km/hr (SEGMENT_70KMHR), four-leg intersections, and intersection density were positively associated with Fatal-Injury collisions. Traffic Analysis Zones with higher Vehicle-Kilometer-Traveled is an indicator of how much a road corridor has been used as well as the traffic volume. The higher the Vehicle-Kilometer-Traveled, the higher the chances for collisions, which explains the increasing Fatal-Injury collisions with increase in log-transformed Vehicle-Kilometer-Traveled. Also, road segments with posted speed limits of 70 km/hr are mostly arterials that connect road users to places that attract trips, such as shopping centres, retail businesses, institutions, and work places. This high volume of road users, resulting from trip attraction, leads to road user conflicts between vehicles, motorists, and pedestrians. Such road user conflicts explain why an increase in the length of road corridors with posted speed limits of 70 km/hr results in the increase in Fatal-Injury collisions. Moreover, 4-leg intersections introduce a lot of conflicts between vehicles, motorists, and pedestrians. Therefore, an increase in four-leg intersections lead to an increase in Fatal-Injury collisions. Similarly, an increase in intersection density implies increase in conflict points, which, in turn, leads to increase in Fatal-Injury collisions.

On the other hand, road segments with posted speed limits of 80 km/hr (SEGMENT_80KMHR), 3-leg intersections, log-transformed collector road length, and proportion of urban holding residential areas are negatively associated with Fatal-Injury collisions. Roadways with posted speed limits of 80 km/hr are highways with some form of median between opposing traffic directions and barriers on edges of roadways, as well as minimal or no pedestrian crossing. These road segments result in reduced vehicle-to-vehicle or vehicle-to-pedestrian conflict and, accordingly, results in reduction of Fatal-Injury collisions. Also, 3-leg intersections have minimal conflict points, leading to their reduction effects on Fatal-Injury collisions. Because collector roads are corridors with posted speed limits of 50 or 40 km/hr, collisions on such corridors hardly involve fatalities or injuries. Thus, an increase in the log-transformed length of collector length has a reduction effect on Fatal-Injury collisions. Lastly, urban holding areas are undeveloped land areas that do not attract vehicular and pedestrian traffic; therefore, the higher the proportion of urban holding residential areas in a Traffic Analysis Zone, the fewer Fatal-Injury collisions are expected.

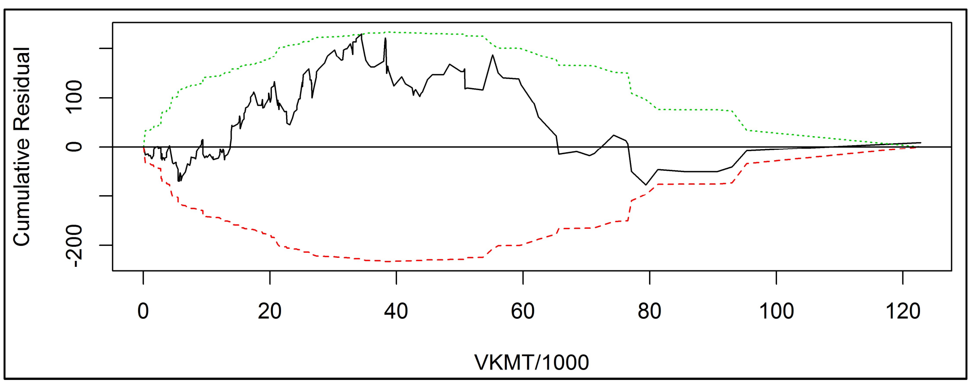


Figure 5.3: Cumulative Residual Plot for Model 10: Fatal-Injury Collisions

Table 5.8: Fatal-Injury Collisions Negative Binomial Regression Model Results

Model 10					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-3.73E+00	5.61E-01	-6.65622	< 0.001	Dispersion Parameter = 2.1548 Standard Error = 0.252 Log-likelihood = -1733.239
logVKMT	5.15E-01	5.99E-02	8.59797	< 0.001	
SEGMENT_80KMHR	-4.04E-04	1.67E-04	-2.41651	0.0157	
SEGMENT_70KMHR	2.35E-04	1.36E-04	1.72792	0.0840	
FOUR_LEG_INTERSECTIONS	3.04E-02	8.74E-03	3.48164	< 0.001	
INTERSECTION_DENSITY	6.71E-01	2.14E-01	3.14138	0.0017	
THREE_LEG_INTERSECTIONS	-1.40E-02	6.53E-03	-2.14826	0.0317	
log1p(COLLECTOR_LENGTH)	-2.60E+00	9.00E-01	-2.88413	0.0039	
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	-2.72E+00	3.74E-01	-7.27061	< 0.001	

Mathematical representation of the selected Fatal-Injury collision model is shown in Equation 5.2. Appendix H provides the mathematical representation of the top 10 Fatal-Injury collision models.

$$FI = N \times \exp(-3.73) \times \exp\left(\begin{array}{l} (\log VKMT \times 0.515) + (SEGMENT\_80KMHR \times -4.04 \times 10^{-4}) + \\ (SEGMENT\_70KMHR \times -2.35 \times 10^{-4}) + (FOUR\_LEG\_INTERSECTIONS \times 0.0304) + \\ (\frac{INTERSECTIONDENSITY}{10000} \times 0.671) + (INTKD \times 0.121) + \\ (THREE\_LEG\_INTERSECTIONS \times -0.014) + \left(\log \frac{COLLECTOR\_LENGTH}{10000} \times -2.6\right) + \\ (URBAN\_HOLDING\_RESIDENTIAL\_AREA\_PROP \times -2.72) \end{array}\right) \quad 5.2$$

Predicted numbers of Fatal-Injury collisions were determined using the estimates from the selected Fatal-Injury collision model. Expected numbers of Fatal-Injury collisions per Traffic Analysis Zone were estimated using Empirical Bayes approach. Figure 5.4 is a map depicting the various hotspots for Fatal-Injury collisions.

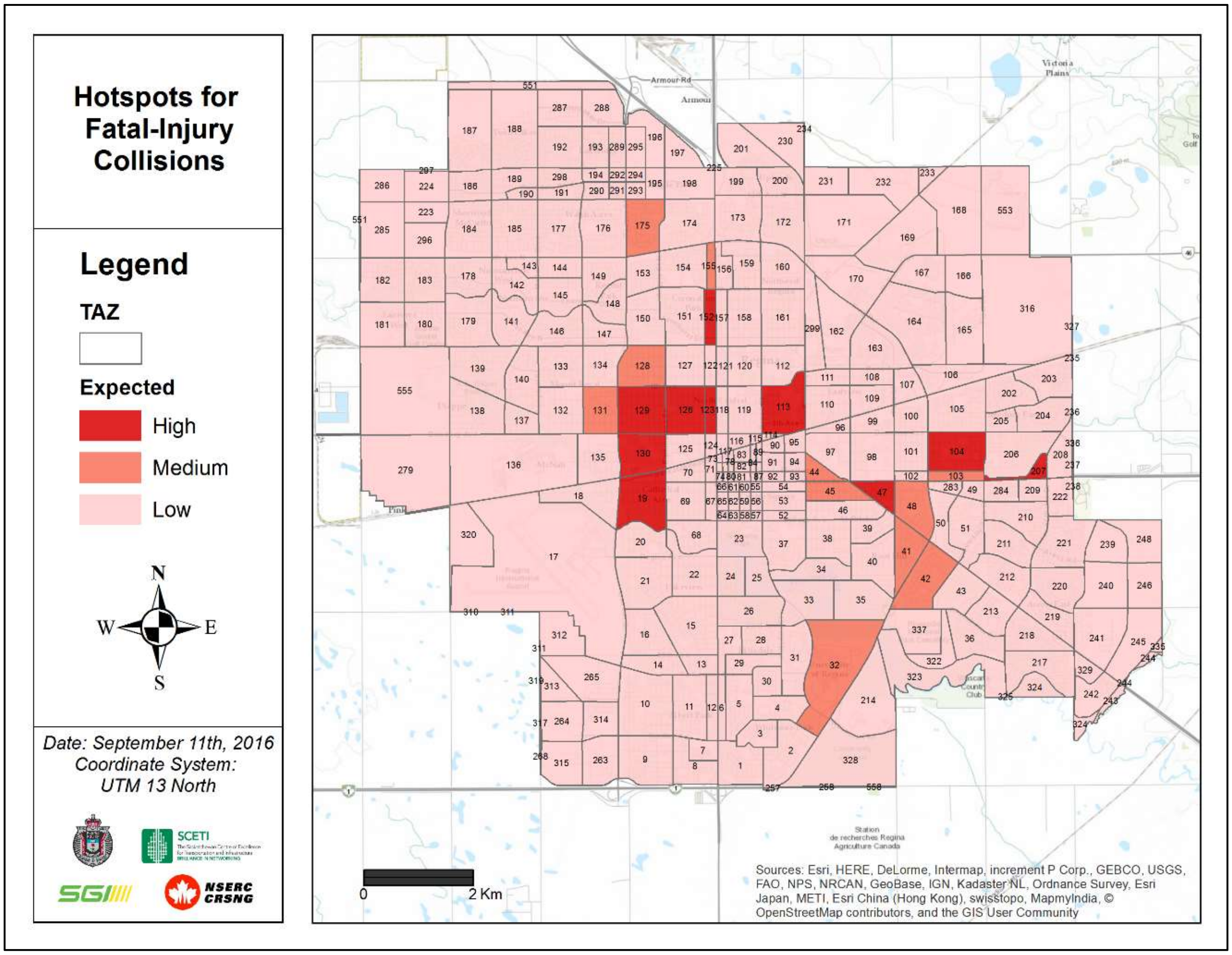


Figure 5.4: Hotspot Map for Expected Numbers of Fatal-Injury Collisions

TAZs were then ranked based on the numbers of expected Fatal-Injury collisions. The top 10 ranked TAZs were grouped as high, TAZs ranked from 11 to 20 were grouped as medium, and TAZs with ranks below 20 were grouped as low. Table 5.9 gives the range of values of expected Fatal-Injury collisions for the three hotspot levels: High, medium, and Low as indicated in the map legend. The numbers of TAZs within those ranges are also presented in Table 5.9.

Table 5.9: Expected Numbers of Fatal-Injury Collisions per Traffic Analysis Zone Legend

Map Legend	Number of collisions	Number of Traffic Analysis Zones
High	72.80 – 103.07	10
Medium	56.80 – 72.74	11
Low	Less than 56.64	241

Table 5.10 provides the numbers of the expected Fatal-Injury collisions for the top 10 hotspots.

Table 5.10: Top 10 Fatal-Injury Collision Hotspots

Rank	Traffic Analysis Zone Number	Number of Expected Fatal-Injury Collisions
1	126	103.0672
2	207	100.5821
3	113	94.71618
4	47	93.74692
5	130	89.57795
6	152	85.10229
7	19	80.67682
8	123	80.51871
9	129	77.67984
10	104	72.80347

5.1.3 Property Damage Only Collision

Property Damage Only Collision Prediction Models showed a similar trend to those of total collisions, which is expected because total collisions are comprised of mostly Property Damage Only collisions (79%). Property Damage Only Collision Prediction Models also showed a slightly better prediction compared to total collisions since the +2 and -2 standard deviations were lower. Table 5.11 is a summary of the goodness-of-fit tests results for the top 10 Property Damage Only collisions. Based on the various goodness-of-fit tests, model 1 showed the best predictive performance and fit to the data and was the selected Property Damage Only collision model. Covariates and their corresponding p-values for the top 10 Property Damage Only collisions models are presented in Table 5.12. Log-transformed Vehicle-Kilometer-Traveled and 4-leg intersections were the most significant predictors. Most predictor variables had p-values less than 0.05. In a few instances, predictors variables had p-values greater than 0.05 but were maintained in their respective models because they improved the predictive performance of models.

Table 5.11: Summary of Result of goodness-of-fit Tests for Property Damage Only Collisions

Model	AIC	BIC	Dispersion Parameter	MSE	MSPE	R2FT		MPB	MAD	RMSE
						Calibration Data	Validation Data			
1	**2351.60**	**2382.78**	**2.01**	**2039.48**	**2065.58**	**0.65**	**0.70**	**-11.88**	**32.95**	**45.16**
2	2355.61	2383.32	1.95	2028.49	2194.97	0.64	0.68	-10.05	32.98	45.04
3	2411.86	2446.49	1.56	2976.07	3479.76	0.60	0.52	-3.89	42.07	54.55
4	2411.93	2436.18	1.51	2352.12	2495.27	0.59	0.59	-6.80	37.05	48.50
5	2352.00	2383.18	2.00	2135.98	2275.16	0.64	0.68	-11.32	34.12	46.22
6	2412.19	2436.44	1.51	2513.64	3068.30	0.58	0.54	-5.57	42.41	50.14
7	2442.39	2470.10	1.32	2877.17	3021.41	0.52	0.47	-6.71	41.83	53.64
8	2396.33	2424.04	1.64	2383.06	2548.11	0.61	0.60	-5.16	39.39	48.82
9	2360.55	2391.72	1.93	2494.53	2993.20	0.62	0.60	-10.72	39.81	49.95
10	2360.57	2391.75	1.93	2458.33	3068.46	0.62	0.60	-10.41	40.20	49.58

Table 5.12: Top 10 Property Damage Only Collisions Models Covariates and their P-Values

Covariate	P-Value
Model 1	
Intercept	< 0.001
logVKMT	< 0.001
I3WP	0.0527
INTKD	< 0.001
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	< 0.001
FOUR_LEG_INTERSECTIONS	< 0.001
SEGMENT_80KMHR	0.0160
LOW_DENSITY_RESIDENTIAL_AREA_PROP	0.0131
Model 2	
Intercept	< 0.001
logVKMT	< 0.001
I3WP	0.0046
INTKD	< 0.001
FOUR_LEG_INTERSECTIONS	0.0111
SEGMENT_80KMHR	0.0227
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	< 0.001
Model 3	
Intercept	< 0.001
logVKMT	< 0.001
FOUR_LEG_INTERSECTIONS	< 0.001
INTERSECTION_DENSITY	< 0.001
SEGMENT_80KMHR	< 0.001
THREE_LEG_INTERSECTIONS	0.1157
LOW_DENSITY_RESIDENTIAL_AREA_PROP	0.1050
POPULATION_18TO24	0.0372
POPULATION_45TO64	0.0724
Model 4	
Intercept	< 0.001
logVKMT	< 0.001
FOUR_LEG_INTERSECTIONS	< 0.001
INTERSECTION_DENSITY	< 0.001
SEGMENT_80KMHR	< 0.001
LOW_DENSITY_RESIDENTIAL_AREA_PROP	0.0186
Model 5	
(Intercept)	< 0.001
logVKMT	< 0.001
I3WP	0.0384
INTKD	< 0.001
FOUR_LEG_INTERSECTIONS	0.0018
SEGMENT_80KMHR	0.0212
COLLECTOR_LENGTH	0.0165
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	< 0.001

Table 5.12 Top 10 Property Damage Only Collisions Models Covariates and their P-Values [Cont'd]

Covariate	P-Value
Model 6	
Intercept	< 0.001
logVKMT	< 0.001
SEGMENT_80KMHR	< 0.001
FOUR_LEG_INTERSECTIONS	< 0.001
INTERSECTION_DENSITY	0.0010
THREE_LEG_INTERSECTIONS	0.0119
Model 7	
Intercept	< 0.001
logVKMT	< 0.001
THREE_LEG_INTERSECTIONS	0.2099
POPULATION_45TO64	0.8139
LOW_DENSITY_RESIDENTIAL_AREA_PROP	0.2126
FOUR_LEG_INTERSECTIONS	< 0.001
INTERSECTION_DENSITY	< 0.001
Model 8	
Intercept	< 0.001
logVKMT	< 0.001
I3WP	0.0329
INTKD	< 0.001
FOUR_LEG_INTERSECTIONS	0.0125
SEGMENT_80KMHR	< 0.001
THREE_LEG_INTERSECTIONS	0.0316
Model 9	
Intercept	< 0.001
logVKMT	< 0.001
SEGMENT_80KMHR	0.0195
FOUR_LEG_INTERSECTIONS	< 0.001
INTERSECTION_DENSITY	0.0244
THREE_LEG_INTERSECTIONS	0.0069
ALKP	0.6872
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	< 0.001
Model 10	
Intercept	< 0.001
logVKMT	< 0.001
SEGMENT_80KMHR	0.0172
FOUR_LEG_INTERSECTIONS	< 0.001
INTERSECTION_DENSITY	0.0238
THREE_LEG_INTERSECTIONS	0.0080
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	< 0.001
YOUNG_DRIVERS*	0.7304

*YOUNG DRIVERS refers to populations aged 1 to 17 and 18 to 24 years.

Figure 5.5, the Cumulative Residual plot for the selected Property Damage Only collisions model, showed a good prediction within the +2 and -2 standard deviations over almost the entire range of Vehicle-Kilometer-Traveled. All 8 variables in the selected Property Damage Only collision model were not highly correlated with each other. Except for intersection road density with a p-value of 0.0527, which was maintained in the model due its improvement in the predictive capability of the model, all other variables had p-values less than 0.05. Table 5.13 is a summary of the Negative Binomial regression model results for the selected model. The summary of Negative Binomial regression model results for the top ten models are presented in Appendix H and their corresponding Cumulative Residual plots are presented in the appendix D.

Log-transformed Vehicle-Kilometer-Traveled (logVKMT), intersection road density (INTKD), and 4-leg intersections were positively associated with Property Damage Only collisions. As explained earlier, increasing Vehicle-Kilometer-Traveled implies higher traffic volume and longer kilometers traveled by vehicles, consequently, increasing the risk for a collision. Therefore, increase in Vehicle-Kilometer-Traveled has a positive effect on Property Damage Only collisions. Also, increasing numbers of intersections per unit road segment length and 4-leg intersections increases conflict points in a road, which increase risks for a collision. As such, increase in intersection road density (INTKD) and 4-leg intersections have an increasing effect on Property Damage Only collisions.

However, proportion of 3-leg intersections (I3WP), proportion of urban holding residential areas, road corridors with posted speed limit of 80 km/hr, and proportion of low density residential areas have a negative association with Property Damage Only collisions. 3-leg intersections have fewer conflict points compared to 4-leg intersections; as a result, the number of Property Damage Only collisions are reduced by higher proportion of 3-leg intersections in a Traffic Analysis Zone. Urban holding residential areas are undeveloped lands and, as such, do not attract trips; therefore, there is less or no exposure, influencing a reduction effect on Property Damage Only collisions. Road corridors with posted speed limits of 80 km/r are high speed corridors that have medians, shoulder lanes, and no or fewer intersections, and, as such, the risk for a Property Damage Only collision is significantly reduced. Moreover, low density residential areas are either already developed neighbourhoods with few people or are developing neighbourhoods. Such places do not attract a lot trips, implying a reduction in traffic volume, which, subsequently, has a reduction effect on Property Damage Only collisions.

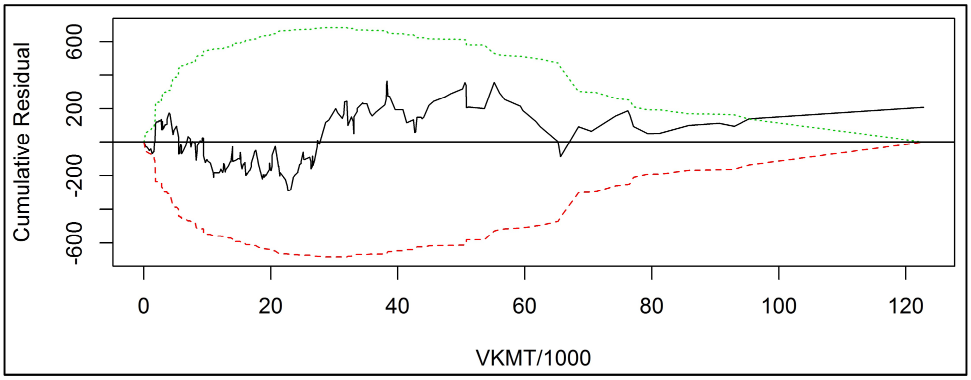


Figure 5.5: Cumulative Residual Plot for Model 8: Property Damage Only Collisions

Table 5.13: Property Damage Only Collisions Negative Binomial Regression Model Results

Covariate	Estimate	Standard Error	Z value	Pr(>|z|)	Model Statistics
Intercept	-1.86E+00	5.20E-01	-3.57808	< 0.001	Dispersion Parameter = 2.0071 Standard Error = 0.210 Log-likelihood = -2333.605
logVKMT	4.43E-01	5.73E-02	7.73169	< 0.001	
I3WP	-3.51E-03	1.81E-03	-1.93722	0.0527	
INTKD	1.24E-01	3.19E-02	3.88672	< 0.001	
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	-2.18E+00	2.75E-01	-7.94025	< 0.001	
FOUR_LEG_INTERSECTIONS	3.78E-02	1.07E-02	3.54489	< 0.001	
SEGMENT_80KMHR	-3.41E-04	1.42E-04	-2.40784	0.0160	
LOW_DENSITY_RESIDENTIAL_AREA_PROP	-4.47E-01	1.80E-01	-2.48110	0.0131	

Equation 5.3 is a mathematical functional form for the selected Property Damage Only collisions model. The mathematical Equations representing the top 10 Property Damage Only collisions models are presented in Appendix H.

$$PDO = N \times \exp(-1.86) \times \exp\begin{pmatrix} (\log VKMT \times 0.443) + (I3WP \times -3.51 \times 10^{-3}) + (INTKD \times 0.124) + \\ (URBAN\_HOLDING\_RESIDENTIAL\_AREA\_PROP \times -2.18) + \\ (FOUR\_LEG\_INTERSECTIONS \times 0.0378) + (SEGMENT\_80KMHR \times -3.41 \times 10^{-4}) + \\ (LOW\_DENSITY\_RESIDENTIAL\_AREA\_PROP \times -0447) \end{pmatrix} \qquad 5.3$$

Predicted numbers of Property Damage Only collisions were estimated using the regression results from the selected model, and expected numbers per Traffic Analysis Zone were determined using Empirical Bayes method. The map for the expected numbers of Property Damage Only collisions hotspot is illustrated in Figure 5.6.

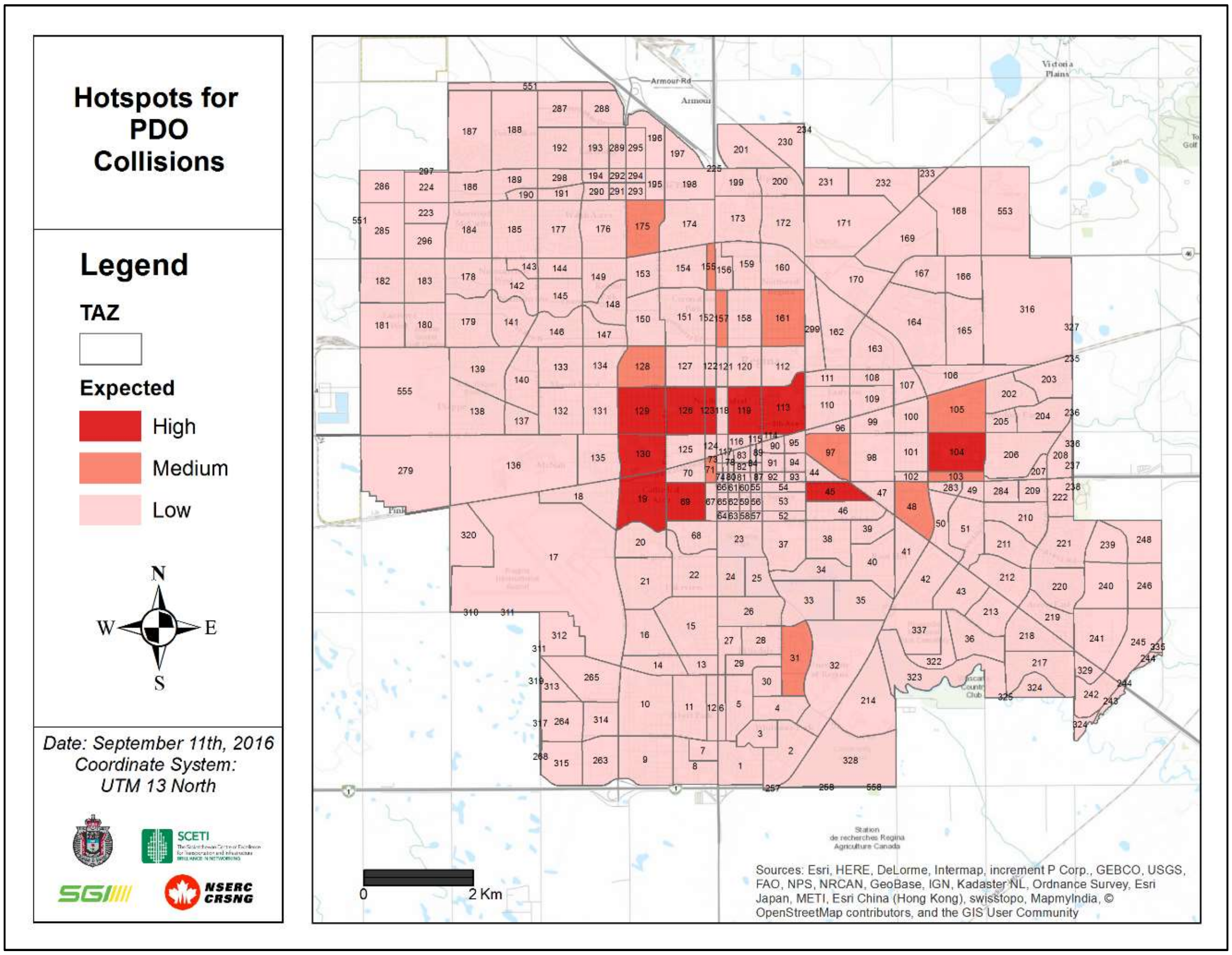


Figure 5.6: Hotspot Map for Expected Numbers of Property Damage Only Collisions

TAZs were then ranked based on the numbers of expected Property Damage Only collisions. Table 5.14 gives the range of values for each hotspot level illustrated in the map as high, medium, and low.

Table 5.14: Expected number of Property Damage Only Collisions per Traffic Analysis Zone Legend

Map Legend	Number of collisions	Number of Traffic Analysis Zones
High	214.75 – 326.36	10
Medium	185.57 – 207.85	11
Low	Less than 185.19	241

Table 5.15 lists the Traffic Analysis Zones with the top 10 highest numbers of expected Property Damage Only collisions.

Table 5.15: Top 10 Property Damage Only Collision Hotspots

Rank	**Traffic Analysis Zone Number**	**Number of Expected Property Damage Only Collisions**
1	129	326.3633
2	126	322.4255
3	113	308.4782
4	119	308.2687
5	123	285.2917
6	19	281.2
7	130	254.9169
8	104	236.2648
9	45	225.7758
10	69	214.7522

5.2 Crime Prediction Model Result

For crime prediction models, there were 42 available variables to be considered in the model development. For each crime occurrence type, several candidate models were created, and the top 6 were selected based on the results of the goodness-of-fit tests. The best out of the top 6 candidate models was then chosen as the selected model. There were fewer predictors for crimes and as such fewer models were created and that explains why 6 candidate models are selected from crime prediction. Models were created for eight different occurrence types: violent, assault, robbery, break and enter, mischief, theft, theft from auto, and theft of auto. Models were not created for sexual assault, arson, and murder because there were very few recorded numbers of incidents, and, as such, the majority of Traffic Analysis Zones had zero values.

This would have caused a zero-bias or zero-inflated in the results of models developed; therefore, for the purpose of avoiding a skewed model, models were not created for those individual crime occurrence types. Similarly, models were not created for aggregated non-violent crimes (break and enter, mischief, theft, theft from auto, and theft of auto) because the aggregated data were sparsely distributed and that introduced a wide variability in the model creation. As a result, individual models were created for the five non-violent crimes listed above. The results of the Negative Binomial regression model for the selected model is presented in this section, and the results of Negative Binomial regression models for the top six models for each crime type are presented in the Appendix H. Cumulative Residual plots for the top six models are also presented in the Appendix D.

5.2.1 Violent Crimes

Violent crimes included arson, assault, murder, robbery, and sexual assault. They each involve some level of violence and, as such, were grouped as violent crimes. Several candidate models were created and the top 6 models were selected based on the results from the goodness-of-fit tests. Table 5.16 presents the summary of the goodness-of-fit test for the top 6 candidate models.
Evident from the results, model 1 provided the best fit for violent crime data and was the selected violent crimes model. Table 5.17 provides a list of the different variables in in the top 6 violent crimes models and their corresponding p-values. As can be seen from Table 5.17, p-values of most covariates were significant with values below 0.05. However, in some models, some variables had p-values greater than 0.05 but were maintained in their respective models because they improved significantly, the predictive capability of models.

Table 5.16: Summary of Result of goodness-of-fit Tests for Violent Crimes

Model	AIC	BIC	Dispersion Parameter	MSE	MSPE	R2FT		MPB	MAD	RMSE
						Calibration Data	Validation Data			
1	**1570.42**	**1596.43**	**1.08**	**5124.25**	**1297.32**	**0.40**	**0.51**	**-9.34**	**18.27**	**71.58**
2	1619.89	1645.90	0.83	7498.67	3028.55	0.20	0.14	5.28	26.43	86.59
3	1601.34	1630.61	0.92	6474.76	2388.56	0.30	0.26	2.97	25.39	80.47
4	1571.56	1597.58	1.07	5409.22	1340.66	0.38	0.51	-9.38	18.46	73.55
5	1617.67	1640.44	0.82	7612.62	3385.12	0.07	0.20	-11.29	23.74	87.25
6	1576.31	1599.07	1.04	5216.27	1163.97	0.35	0.50	-9.37	18.12	72.22

Table 5.17: Top 10 Violent Crimes Models Covariates and their P-Values

Covariate	P-Value
Model 1	
(Intercept)	0.19496
log1p(COMMERCIAL_AREA)	<0.001
POPULATION_DENSITY	0.0022
POPULATION_25TO44	<0.001
POPULATION_45TO64	<0.001
POPULATION_18TO24	0.1385
RETAIL_SPACE	0.0799
Model 2	
(Intercept)	0.0033
POPULATION_DENSITY	<0.001
POPULATION_18TO24	0.7963
POPULATION_25TO44	0.2772
RETAIL_SPACE	<0.001
LAND_USE_PER_TAZ	<0.001
HIGH_DENSITY_RESIDENTIAL_AREA	0.0586
Model 3	
(Intercept)	<0.001
POPULATION_18TO24	0.0581
POPULATION_25TO44	<0.001
TOT_POP	<0.001
POPULATION_DENSITY	<0.001
LAND_USE_PER_TAZ	<0.001
OFFICE_SPACE	0.0048
RETAIL_SPACE	<0.001
Model 4	
(Intercept)	0.1074
log1p(COMMERCIAL_AREA)	<0.001
POPULATION_DENSITY	<0.001
LOW_DENSITY_RESIDENTIAL_AREA	0.2283
RETAIL_SPACE	0.0766
POPULATION_25TO44	<0.001
POPULATION_45TO64	<0.001
Model 5	
(Intercept)	0.0032
log1p(COMMERCIAL_AREA)	<0.001
POPULATION_DENSITY	<0.001
INDUSTRY_SPACE	0.0465
LOW_DENSITY_RESIDENTIAL_AREA	<0.001
OFFICE_SPACE	0.6935

Table 5.17: Top 10 Violent Crimes Models Covariates and their P-Values [cont'd]

Covariate	P-Value
Model 6	
(Intercept)	0.9775
log1p(COMMERCIAL_AREA)	<0.001
POPULATION_18TO24	0.2107
POPULATION_25TO44	<0.001
POPULATION_45TO64	<0.001
RETAIL_SPACE	0.1912

Figure 5.7, a Cumulative Residual plot for the selected violent crimes model and it shows a good fit until the 700 mark on the x-axis with residuals within the +2 and -2 standard deviations, which are very close to the zero line. Table 5.18 gives the summary of the results of the Negative Binomial regression model for the selected violent crime model. As can be seen from the regression results, log-transformed commercial area, population density, population aged 24 to 44, population aged 18 to 24, and retail space are positively associated with violent crimes, whereas population aged 45 to 64 is negatively associated with violent crimes.

Commercial areas as well as retail spaces have high attraction for shopping, personal, business, and leisure trips. The high frequency of trips often leads to an increase in human and vehicular traffic. The traffic, in turn, makes such places targets for some criminal activities, such as assault, robbery, and sexual assault and, hence, the positive association with violent crimes. Also, increasing population density increases the demand for shopping malls, retail stores, and commercial areas, and, as previously explained, such attractions have high associations with violent crimes. Moreover, the most active population age groups, 18 to 25 and 25 to 44 years, also have a positive effect on violent crimes because such people are more active and involved in many activities and vices. Therefore, even though these groups constitute the majority of the age group for the workforce of any economy, it is not surprising that they contribute significantly to violent crimes. Other findings by researchers claimed that the population aged 18-45 are the most unemployed and, as such, tend to engage in crimes to sustain themselves; however, unemployment data by Traffic Analysis Zone was not available to support such findings.

On the other hand, the number of residents between the age group of 45 and 64 have a negative effect on the number of violent crimes. It is expected that between the age of 45 and 64, most residents would are settled with secure jobs and are established. Moreover, such people are usually stable economically and are not physically active, thus, explaining the reduction effect population aged 45 to 64 years have on violent crimes.

Similarly, the other two violent crimes, assault and robbery crimes, had similar results with the same predictors.

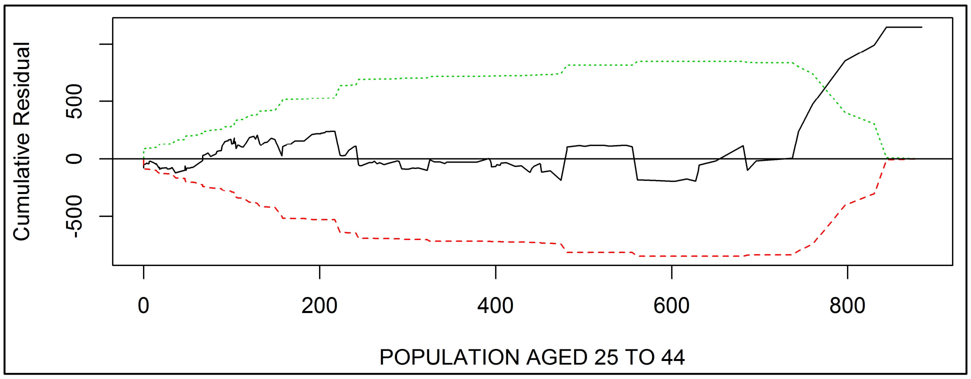


Figure 5.7: Cumulative Residual Plot for Model 5: Five-Violent Crimes

Table 5.18: Violent Crimes Negative Binomial Regression Model Results

Covariate	Estimate	Standard Error	Z value	Pr(>|z|)	Model Statistics
(Intercept)	-1.89E-01	1.46E-01	-1.296	0.19496	Dispersion Parameter = 1.080 Standard Error = 0.118 Log-likelihood = -1554.416
log1p(COMMERCIAL_AREA)	5.74E-01	9.58E-02	5.98939	<0.001	
POPULATION_DENSITY	1.56E-04	5.07E-05	3.06586	0.0022	
POPULATION_25TO44	8.46E-03	1.40E-03	6.06211	<0.001	
POPULATION_45TO64	-7.56E-03	1.35E-03	-5.5864	<0.001	
POPULATION_18TO24	4.80E-03	3.24E-03	1.4813	0.1385	
RETAIL_SPACE	1.12E-01	6.37E-02	1.75139	0.0799	

Equation 5.4 is a mathematical representation of the selected violent crime model. Mathematical equations representing all top six violent crime models are presented in Appendix H.

$$VIOLENT\_CRIMES = N \times \exp(-0.189) \times \exp\left(\begin{array}{l} \left(\log(\frac{COMMERCIAL\_AREA}{10000}) \times 0.574 \right) + (POPULATION\_DENSITY \times 1.56 \times 10^{-4}) + \\ (POPULATION\_25TO44 \times 8.46 \times 10^{-4}) + (POPULATION\_45TO64 \times -7.56 \times 10^{-3}) + \\ (POPULATION\_18TO24 \times 4.8 \times 10^{-3}) + (\frac{RETAIL\_SPACE}{10000} \times 0.112) \end{array} \right) \qquad 5.4$$

Violent crimes were then predicted using estimates from the selected violent crime model. Empirical Bayes method was applied to the predicted numbers of violent crimes, and expected numbers were calculated. Figure 5.8 is a hotspot map depicting the expected number of violent crimes by Traffic Analysis Zones for the City of Regina. Using the Empirical Bayes method for calculations, this expected number represents the predicted numbers that have been compensated for with the observed number. TAZs were then ranked based on the numbers of violent expected crimes. The top 10 ranked TAZs were grouped as high, top 11 to 20 ranked TAZs were grouped as medium and TAZs ranked 21 and below were grouped as; low, as illustrated in the map in Figure 5.8. Table 5.19 gives the range of values of the expected violent crimes for the three hotspot groups.

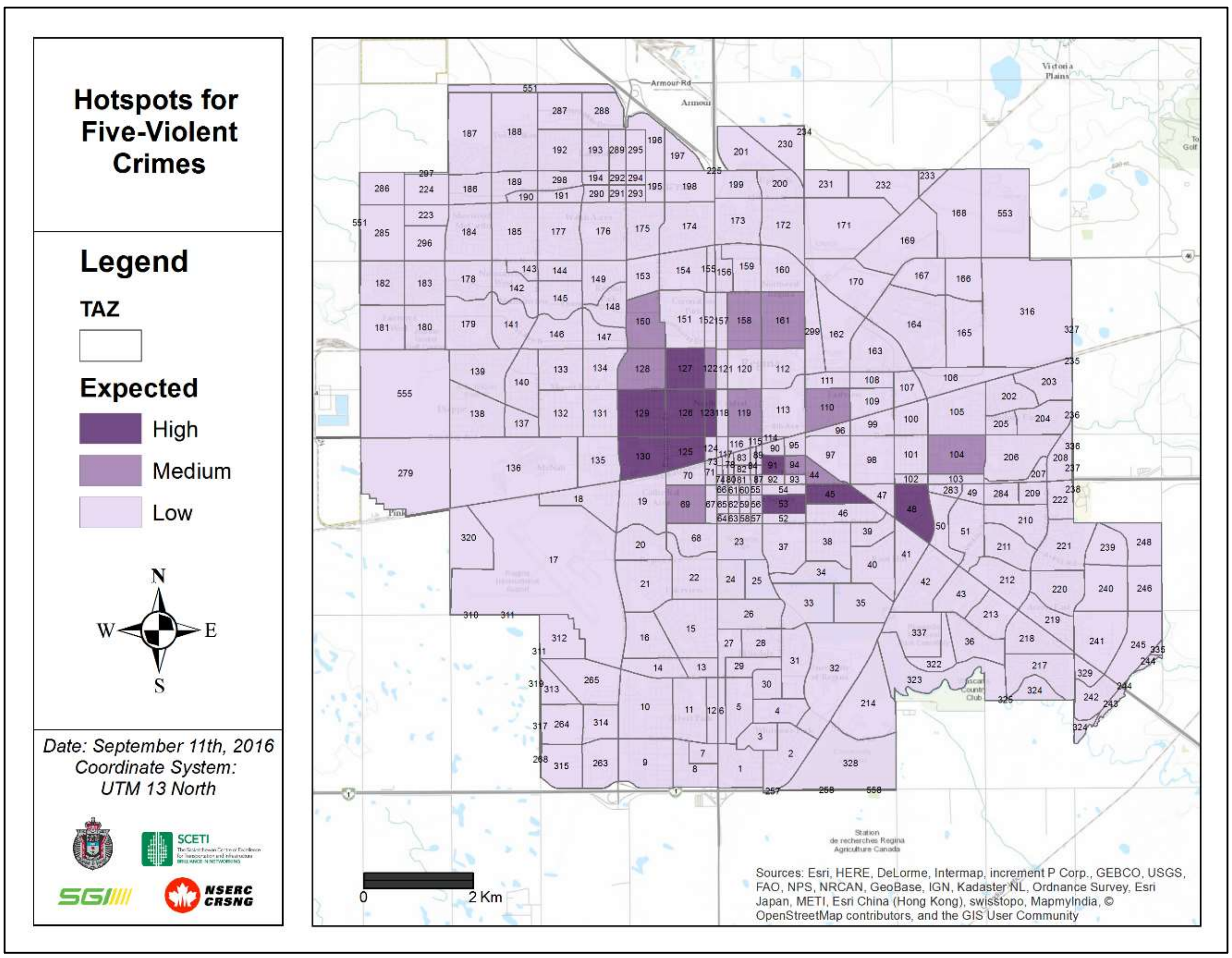


Figure 5.8: Hotspot Map for Expected Numbers of Violent Crimes

Table 5.19: Expected number of Five-Violent Crimes per Traffic Analysis Zone Legend

Map Legend	Number of Violent Crimes	Number of Traffic Analysis Zones
High	136.98 – 942.30	10
Medium	90.2 – 131.98	10
Low	Less than 86.49	242

Table 5.20 gives the numbers of expected violent for the top ten violent crimes Traffic Analysis Zones hotspots.

Table 5.20: Top 10 Five-Violent Crime Hotspots

Rank	Traffic Analysis Zone Number	Number of Expected Violent Crimes
1	126	942.30
2	129	480.03
3	127	392.08
4	53	335.22
5	123	311.26
6	125	197.52
7	130	154.89
8	48	145.88
9	91	145.68
10	45	136.99

5.2.2 Assault Crimes

The results for the goodness-of-fit test for the top 6 assault crimes models are presented in Table 5.21. Based on the results of the various goodness-of-fit tests, model 1 provided the best predictive capability and was the selected model to predict assault crimes. Table 5.22 provides the sets of predictor variables in the top 6 assault crimes models and their corresponding p-values. Most of these predictor variables were significant with p-values less than 0.05. There were a few variables that had p-values greater than 0.05 but were maintained in their respective models because they improved the predictive performance of their models. That explains why some models may be variables with p-values greater than 0.05.

Table 5.21: Summary of Result of goodness-of-fit Tests for Assault Crimes

Model	AIC	BIC	Dispersion Parameter	MSE	MSPE	R2FT		MPB	MAD	RMSE
						Calibration Data	Validation Data			
1	**1456.75**	**1482.77**	**1.08**	**2577.28**	**1060.18**	**0.42**	**0.54**	**5.55**	**15.21**	**50.77**
2	1513.19	1539.21	0.83	4637.09	1582.70	0.09	0.14	5.30	20.34	68.10
3	1516.02	1545.29	0.92	4132.23	1655.11	0.15	0.05	5.56	21.00	64.28
4	1464.02	1490.04	1.07	2574.07	830.96	0.48	0.35	3.50	16.32	50.74
5	1479.42	1505.44	0.82	3976.40	608.27	0.24	0.39	5.24	15.68	63.06
6	1458.76	1481.52	1.04	2839.63	1782.58	0.37	0.52	6.89	16.21	53.29

Table 5.22: Top 10 Assault Crimes Models Covariates and their P-Values

Covariate	P-Value
Model 1	
(Intercept)	<0.001
log1p(COMMERCIAL_AREA)	<0.001
POPULATION_DENSITY	0.0261
POPULATION_25TO44	<0.001
POPULATION_45TO64	<0.001
POPULATION_18TO24	0.3545
RETAIL_SPACE	<0.001
Model 2	
(Intercept)	0.0290
Residential_Proportion	<0.001
RETAIL_SPACE_PROP	<0.001
log1p(LOW_DENSITY_RESIDENTIAL_AREA)	0.2556
LAND_USE_PER_TAZ	<0.001
POPULATION_25TO44	<0.001
COMMERCIAL_AREA	0.0044
Model 3	
(Intercept)	<0.001
POPULATION_25TO44	<0.001
INDUSTRY_SPACE	0.0561
log1p(POPULATION_DENSITY)	0.0041
RETAIL_SPACE_PROP	<0.001
log1p(LOW_DENSITY_RESIDENTIAL_AREA)	0.4824
LAND_USE_PER_TAZ	<0.001
COMMERCIAL_AREA	<0.001
Model 4	
(Intercept)	0.0048
POPULATION_DENSITY	<0.001
RETAIL_SPACE	<0.001
POPULATION_25TO44	<0.001
POPULATION_45TO64	<0.001
LAND_USE_PER_TAZ	<0.001
Residential_Area	<0.001
Model 5	
(Intercept)	<0.001
log1p(COMMERCIAL_AREA)	<0.001
INDUSTRY_SPACE	0.0244
OFFICE_SPACE	0.1559
POPULATION_25TO44	<0.001
POPULATION_65_PLUS	0.0193
Log_Population_Density	0.2422

Table 5.22: Top 10 Assault Crimes Models Covariates and their P-Values [Cont'd]

Covariate	P-Value
Model 6	
(Intercept)	<0.001
log1p(COMMERCIAL_AREA)	<0.001
POPULATION_18TO24	0.4319
POPULATION_25TO44	<0.001
POPULATION_45TO64	<0.001
RETAIL_SPACE	0.0060

The results of the Negative Binomial regression model for predicting assault crimes are summarised in Table 5.23. Cumulative Residual plot for the selected assault crimes model is shown in Figure 5.9. The Cumulative Residual plot depicts a good predictive model with residuals very close to the zero line until the 9 mark on the x-axis (log-transformed commercial areas). As mentioned earlier, the predictors for assault crimes were the same for violent crimes, and the effects of the predictors for assault crimes were similar to that of violent crimes.

Log transformed commercial area, population density, population aged 25 to 44, population aged 18 to 24, and retail space were positively associated with assault crimes. As previously explained, commercial and retail areas are sites that attract many people for various reasons: shopping, business, personal, and pleasure purposes. Such areas are targets for assault crime offenders.

However, an increase in population aged 45 to 65 results in reduction in assault crimes. This can be explained by the fact that such age groups constitute many stable residents, and, as such, they do not engage in criminal activities. Moreover, according to the Regina Census Metropolitan Area data, residents aged 45 to 64 have the highest median individual income (City of Regina, 2011), which could also be a contributing factor for the age group not engaging in assault crimes.

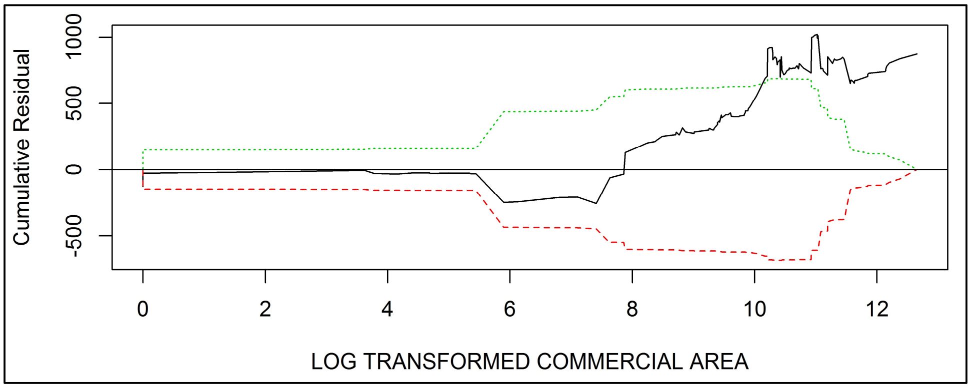


Figure 5.9: Cumulative Residual Plot for Model 5: Assault Crimes

Assault crimes were predicted using estimates from the selected assault crime Negative Binomial regression model results and expected numbers of assault crimes, were estimated using the Empirical Bayes approach.

Table 5.23: Assault Crimes Negative Binomial Regression Model Results

Covariate	Estimate	Standard Error	Z value	Pr(>|z|)	Model Statistics
(Intercept)	-7.80E-01	1.62E-01	-4.823255	<0.001	Dispersion Parameter = 1.1113
log1p(COMMERCIAL_AREA)	1.21E-01	1.81E-02	6.67243	<0.001	Standard Error = 0.127
POPULATION_DENSITY	1.16E-04	5.19E-05	2.224809	0.0261	Log-likelihood = -1440.754
POPULATION_25TO44	8.47E-03	1.39E-03	6.08929	<0.001	
POPULATION_45TO64	-6.68E-03	1.35E-03	-4.937962	<0.001	
POPULATION_18TO24	2.99E-03	3.23E-03	0.925903	0.3545	
RETAIL_SPACE	1.72E-05	5.70E-06	3.019782	<0.001	

Equation 5.5 is a mathematical representation of the selected assault crime model. The mathematical representation of the top six assault crimes are presented in Appendix H.

$$ASSAULT\_CRIMES = N \times \exp(-0.78) \times \exp\begin{pmatrix} (\log COMMERCIAL\_AREA \times 0.121) + (POPULATION\_DENSITY \times 1.16 \times 10^{-4}) + \\ (POPULATION\_25TO44 \times 8.47 \times 10^{-3}) + (POPULATION\_45TO64 \times -6.68 \times 10^{-3}) + \\ (POPULATION\_18TO24 \times 2.99 \times 10^{-3}) + (RETAIL\_SPACE \times 1.72 \times 10^{-5}) \end{pmatrix} \quad 5.5$$

Figure 5.10 is a hotspot maps for the expected numbers of assault crimes per Traffic Analysis Zone. As illustrated in the map, TAZs with the top ten highest numbers of expected assault crimes were labelled as high. TAZs with the top 11 to 20 expected numbers of assault crimes were labelled as medium and the remaining TAZs labelled as low. Table 5.24 gives the range of values of expected assault crimes defined as high, medium, and low in the map.

TAZ numbers and the numbers of expected assault crimes for these top 10 ranked TAZs hotspots are presented in Table 5.25.

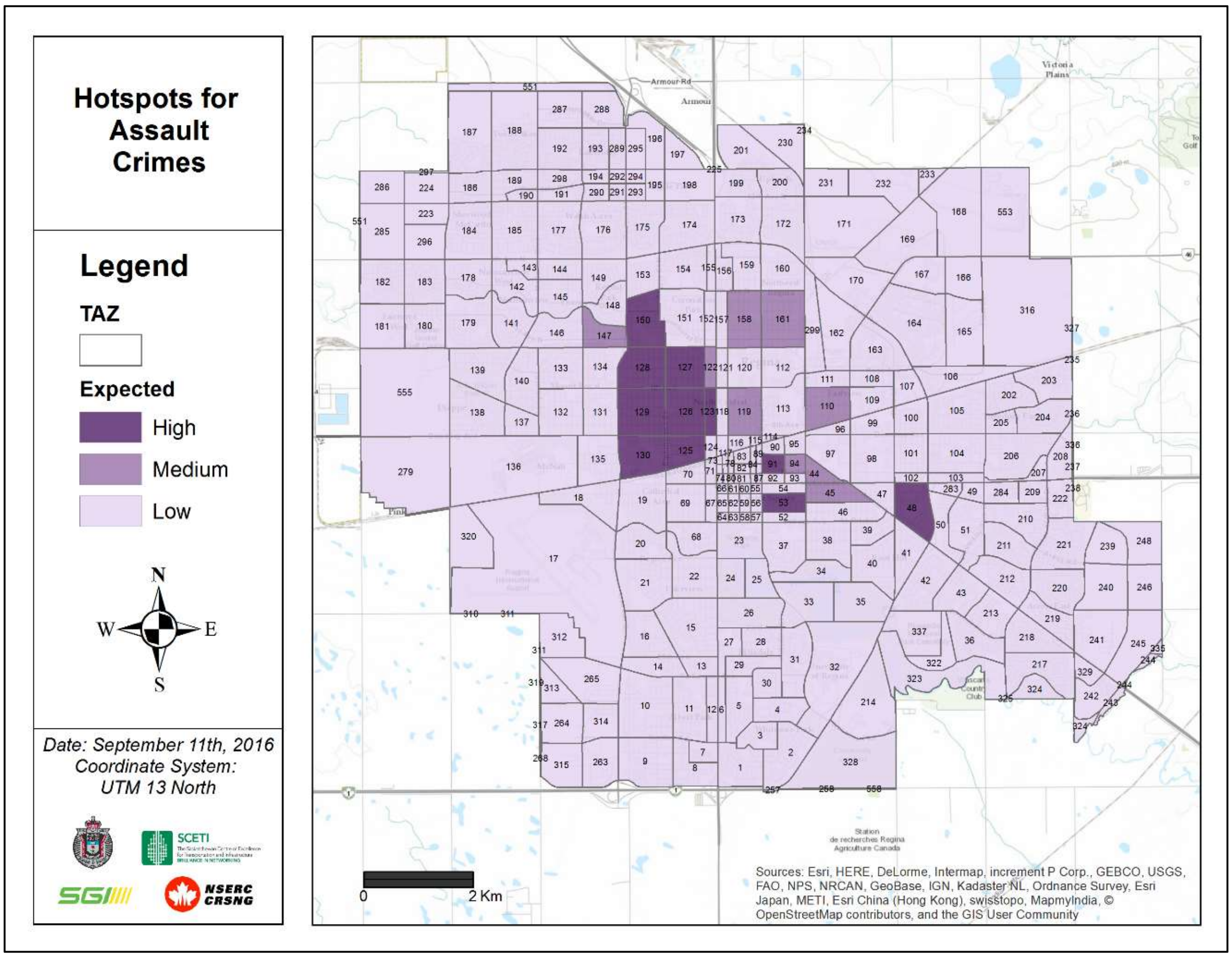


Figure 5.10: Hotspot Map for Expected Number of Assault Crimes

Table 5.24: Expected number of Assault Crimes per Traffic Analysis Zone Legend

Map Legend	Number of Assault Crimes	Number of Traffic Analysis Zones
High	109.17 – 683.230	10
Medium	73.60 – 99.00	10
Low	Less than 73.03	242

Table 5.25: Top 10 Assault Crimes Hotspots

Rank	Traffic Analysis Zone Number	Number of Expected Assault Crimes
1	126	683.23
2	129	335.51
3	127	321.35
4	53	234.67
5	123	228.56
6	125	160.11
7	48	116.57
8	130	113.63
9	91	111.25
10	128	109.17

5.2.3 Robbery Crimes

Several models were created to predict robbery crimes, and the best 6 models were selected based on their goodness-of-fit test results. Table 5.26 is a summary of the goodness-of-fit tests results for the best 6 models. Model 1 had the overall best predictive performance among the six candidate models and was chosen as the selected model and was the selected Robbery crimes model. The top six models had different combinations of predictor variables or covariates. The different covariates had different significance in the models they were kept in. Table 5.27 presents the different sets of variables in the top six Robbery crimes models and their corresponding p-values. Most of these covariates had p-values below 0.05 except in few cases. Although some variables had p-values greater than 0.05, they were maintained in their respective models because they improved significantly, the predictive performance of their respective models.

Table 5.26: Summary of Result of goodness-of-fit Tests for Robbery Crimes

Model	AIC	BIC	Dispersion Parameter	MSE	MSPE	R2FT		MPB	MAD	RMSE
						Calibration Data	Validation Data			
1	**1458.76**	**1481.52**	**1.09**	**92.31**	**26.23**	**0.41**	**0.66**	**-0.04**	**2.81**	**9.61**
2	1513.19	1539.21	0.90	134.29	73.83	0.25	0.51	-0.13	3.49	11.59
3	1516.02	1545.29	0.68	300.03	145.08	-0.34	0.07	0.60	4.99	17.32
4	1464.02	1490.04	0.94	119.62	84.02	0.34	0.31	0.03	4.18	10.94
5	1479.42	1505.44	0.97	123.75	62.65	0.28	0.47	0.15	3.71	11.12
6	1456.75	1482.77	1.20	7730.43	42.84	-1.24	0.61	0.14	3.23	87.92

Table 5.27: Top 10 Robbery Crimes Models Covariates and their P-Values

Covariate	**P-Value**
Model 1	
(Intercept)	<0.001
log1p(COMMERCIAL_AREA)	<0.001
POPULATION_18TO24	0.2877
POPULATION_25TO44	<0.001
POPULATION_45TO64	<0.001
RETAIL_SPACE	0.0275
Covariate	**P-Value**
Model 2	
(Intercept)	<0.001
Residential_Proportion	0.1492
log1p(COMMERCIAL_AREA)	<0.001
POPULATION_18TO24	0.4063
POPULATION_25TO44	0.0031
INDUSTRY_SPACE	0.5664
OFFICE_SPACE	0.2555
Model 3	
(Intercept)	<0.001
POPULATION_25TO44	<0.001
INDUSTRY_SPACE	0.3698
log1p(POPULATION_DENSITY)	0.0242
RETAIL_SPACE_PROP	0.0063
log1p(LOW_DENSITY_RESIDENTIAL_AREA)	0.6206
LAND_USE_PER_TAZ	0.1776
COMMERCIAL_AREA	<0.001
Model 4	
(Intercept)	<0.001
POPULATION_DENSITY	0.0034
RETAIL_SPACE	<0.001
POPULATION_25TO44	<0.001
POPULATION_45TO64	<0.001
LAND_USE_PER_TAZ	0.1845
Residential_Area	0.0040
Model 5	
(Intercept)	<0.001
log1p(COMMERCIAL_AREA)	<0.001
INDUSTRY_SPACE	0.1228
OFFICE_SPACE	0.1451
POPULATION_25TO44	<0.001
POPULATION_65_PLUS	0.002978
Log_Population_Density	0.238343

Table 5.27 Top 10 Robbery Crimes Models Covariates and their P-Values [Cont'd]

Model 6	
Covariate	**P-Value**
(Intercept)	<0.001
log1p(COMMERCIAL_AREA)	<0.001
POPULATION_DENSITY	0.0021
POPULATION_25TO44	<0.001
POPULATION_45TO64	<0.001
log1p(RETAIL_SPACE)	0.0022
Residential_Proportion	0.0078

Table 5.28 is a summary of the results of the selected Negative Binomial regression model for robbery crimes. Cumulative Residual plot for this selected model is shown in Figure 5.11, which shows a good predictive model with residuals close to the zero line and is almost parallel to the zero line across most of the range of commercial area. The cumulative residual converges at the zero line, implying the number of predicted robbery crimes are comparable to the actual observed number of robbery crimes. Negative Binomial model estimates and Cumulative Residual plots for the top 6 models are presented in the appendices H and D respectively.

Increasing log-transformed commercial area, population aged 18 to 24, population aged 25 to 44, and retail space results in an increase in robbery crimes. Intuitively, shopping malls, jewelry stores, and general businesses have numerous items of value and are, therefore, target areas for robberies. As such, commercial areas and retail spaces are places that are positively associated to robbery crimes. Also, the National Household Survey provided by the City of Regina showed that residents within the ages 18 to 44 had the lowest median individual income (City of Regina, 2011). Due to unemployment or lower earnings for people within that age group, this lower median income could be an explanation for the increase in robbery crimes with an increase in those age groups. On the other hand, increases in population aged 45 to 64 reduces robbery crimes, due to the fact that the majority of residents within that age group are economically stable and settled. Accordingly, the majority of such people would not engage in robbery crimes.

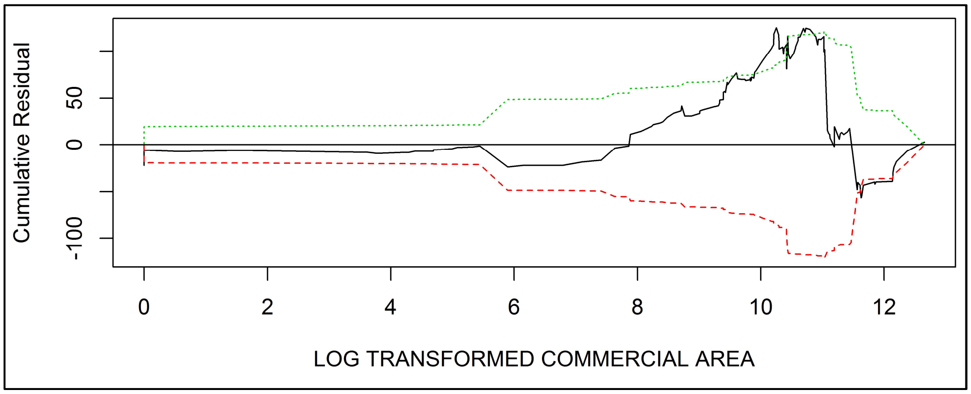


Figure 5.11: Cumulative Residual Plot for Model 6: Robbery Crimes

Table 5.28: Robbery Crimes Negative Binomial Regression Model Results

Covariate	Estimate	Standard Error	Z value	Pr(>|z|)	Model Statistics
(Intercept)	-2.22E+00	2.09E-01	-10.6059	<0.001	Dispersion Parameter = 1.093 Standard Error = 0.162 Log-likelihood = -873.527
log1p(COMMERCIAL_AREA)	1.69E-01	2.15E-02	7.863565	<0.001	
POPULATION_18TO24	3.85E-03	3.63E-03	1.063128	0.2877	
POPULATION_25TO44	7.58E-03	1.57E-03	4.818846	<0.001	
POPULATION_45TO64	-7.35E-03	1.60E-03	-4.60895	<0.001	
RETAIL_SPACE	1.33E-05	6.01E-06	2.204654	0.0275	

Mathematical equation representing the selected robbery crime model is shown in Equation 5.6. Equations representing the top 6 Robbery crimes models are presented in Appendix H.

$$ROBBERY\_CRIMES = N \times \exp(-2.22) \times \exp\begin{pmatrix} (\log COMMERCIAL\_AREA \times 0.169) + (POPULATION\_18TO24 \times 3.85 \times 10^{-3}) + \\ (POPULATION\_25TO44 \times 7.58 \times 10^{-3}) + (POPULATION\_45TO64 \times -7.35 \times 10^{-3}) + \\ (RETAIL\_SPACE \times 1.33 \times 10^{-5}) \end{pmatrix} \quad 5.6$$

Figure 5.12 is a hotspot map of expected numbers of robbery crimes in Traffic Analysis Zone hotspots. Traffic Analysis Zones were ranked based on the expected number of robbery crimes. The top ten ranked Traffic Analysis Zones were assigned as high hotspots; the ranked 11 to 20 were assigned as medium hotspot; and the remaining Traffic Analysis Zones were defined as low hotspots. Table 5.29 gives statistics for the three levels of hotspots identified in the map, and Table 5.30 is a list of the top 10 ranked Traffic Analysis Zones and the number of expected robbery crimes associated with those TAZs.

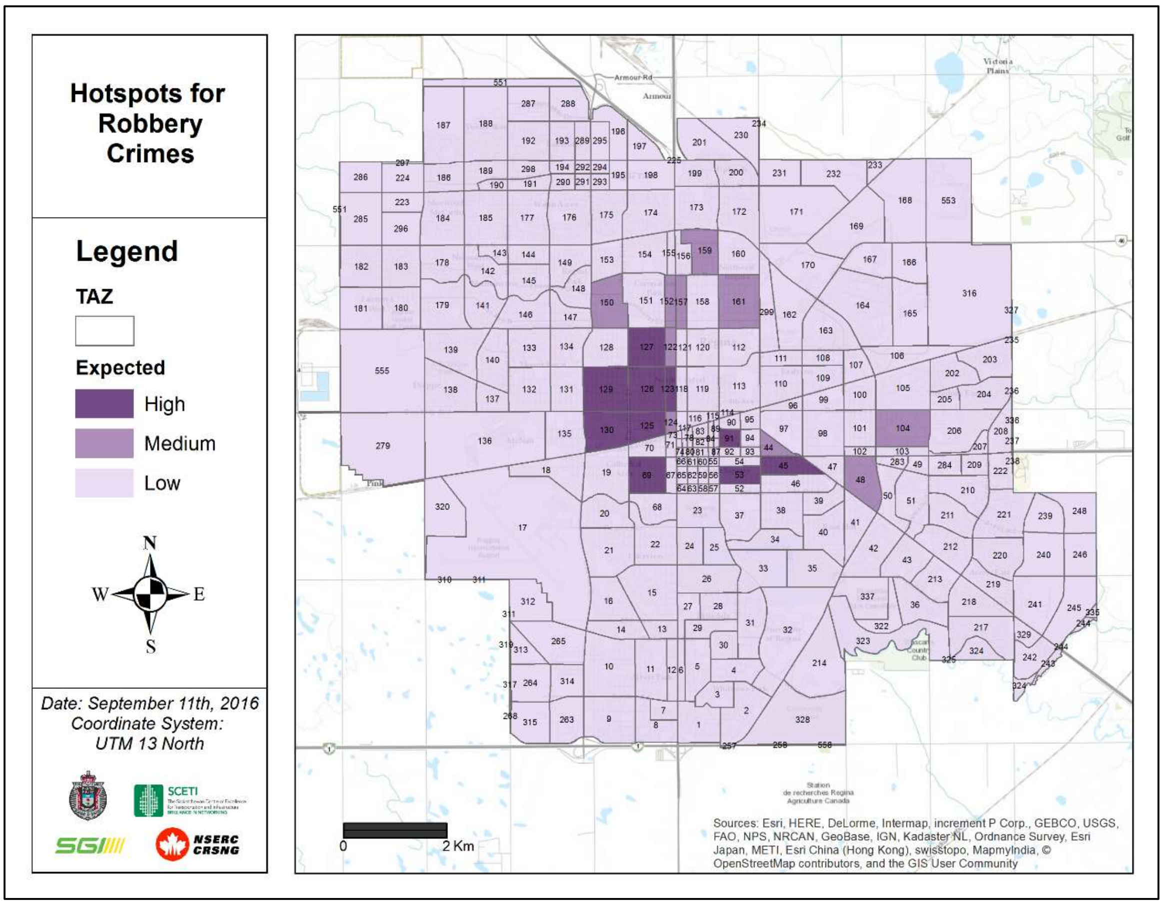


Figure 5.12: Hotspot Map for Expected Number of Robbery Crimes

Table 5.29: Expected number of Robbery Crimes per Traffic Analysis Zone Legend

Map Legend	Number of Robbery Crimes	Number of Traffic Analysis Zones
High	21.97 – 139.30	10
Medium	14.48 – 21.96	10
Low	Less than 14.06	242

Table 5.30: Top 10 Robbery Crimes Hotspots

Rank	Traffic Analysis Zone Number	Number of Expected Robbery Crimes
1	126	139.30
2	129	92.57
3	123	57.08
4	127	33.22
5	53	32.70
6	125	30.89
7	91	24.30
8	69	23.21
9	45	22.46
10	130	21.97

5.2.4 Break and Enter Crimes

Several models were created to predict break and enter crimes, and the results of the goodness-of-fit tests from the best 6 models are summarized in Table 5.31. Model 1 had the overall best predictive performance among the candidate models and was chosen as the selected final model. The top six Break and Enter crimes models each had a set of predictor variables used in prediction. Table 5.32 presents the various variables used in each of the top six Break and Enter crimes models. The p-values of these predictor variables are also presented in Table 5.32. As discussed earlier, in prediction models, some predictor variables may have p-values greater than 0.05 but may significantly improve the predictive capability of models. As such, some variables have p-values greater than 0.05 but were still maintained in their respective models. Most of these variables however, had p-values less than 0.05.

Table 5.31: Summary of Result of goodness-of-fit Tests for Break and Enter Crimes

Model	AIC	BIC	Dispersion Parameter	MSE	MSPE	R2FT		MPB	MAD	RMSE
						Calibration Data	Validation Data			
1	**1471.11**	**1497.12**	**1.80**	**533.565**	**380.36**	**0.536192324**	**0.628355091**	**0.768**	**12.3**	**23.10**
2	1493.57	1519.59	1.58	678.886	483.21	0.439879654	0.476318055	-0.274	13.47	26.06
3	1486.32	1512.34	1.66	535.201	404.66	0.540208326	0.552966359	-0.53	12.59	23.13
4	1478.57	1501.34	1.65	951.823	1034781.96	0.346108036	-14.29918665	127.4	138.2	30.85
5	1483.64	1509.66	1.68	644.637	468.58	0.470685065	0.542284723	-0.333	13.18	25.39
6	1482.02	1504.79	1.67	652.708	460.86	0.464911007	0.553636389	-0.259	13.11	25.55

Table 5.32: Top 10 Break and Enter Crimes Models Covariates and their P-Values

Covariate	P-Value
Model 1	
(Intercept)	0.7909
log1p(COMMERCIAL_AREA)	0.0062
POPULATION_DENSITY	0.0245
LOW_DENSITY_RESIDENTIAL_AREA	<0.001
RETAIL_SPACE	<0.001
POPULATION_25TO44	<0.001
POPULATION_45TO64	<0.001
Model 2	
(Intercept)	0.0104
Residential_Proportion	0.0332
log1p(COMMERCIAL_AREA)	0.0014
POPULATION_DENSITY	0.6581
POPULATION_25TO44	<0.001
POPULATION_45TO64	<0.001
OFFICE_SPACE	0.9935
Model 3	
(Intercept)	0.2270
LAND_USE_PER_TAZ	<0.001
POPULATION_25TO44	<0.001
POPULATION_DENSITY	0.0263
POPULATION_45TO64	<0.001
COMMERCIAL_AREA	0.0127
LOW_DENSITY_RESIDENTIAL_AREA	0.0228
Model 4	
(Intercept)	<0.001
log1p(COMMERCIAL_AREA)	<0.001
POPULATION_DENSITY	<0.001
INDUSTRY_SPACE	<0.001
LOW_DENSITY_RESIDENTIAL_AREA	<0.001
OFFICE_SPACE	0.6516
Model 5	
(Intercept)	0.3778
log1p(COMMERCIAL_AREA)	0.0071
POPULATION_DENSITY	0.5352
POPULATION_25TO44	<0.001
POPULATION_45TO64	<0.001
POPULATION_18TO24	0.2376
RETAIL_SPACE	<0.001

Table 5.32 Top 10 Break and Enter Crimes Models Covariates and their P-Values [Cont'd]

Covariate	P-Value
Model 6	
(Intercept)	0.2822
log1p(COMMERCIAL_AREA)	0.0018
POPULATION_18TO24	0.2558
POPULATION_25TO44	<0.001
POPULATION_45TO64	<0.001
RETAIL_SPACE	<0.001

The Cumulative Residual plot for the selected robbery crimes model, shown in Figure 5.13, displays residuals within the +2 and -2 standard deviations and consistency around the zero line over the range of population aged 25 to 44. The results from the selected Negative Binomial regression model are shown in Table 5.33. Log-transformed commercial area, population density, low density residential area, retail space, and population aged 25 to 44 had positive associations to Break and Enter crimes.

Commercial areas and retail spaces are locations for businesses with very valuable items; therefore, it is not surprising that such places are often targeted for Break and Enter. Places with high population density also attract lots of commercial businesses, therefore, increasing Break and Enter crimes. Low density residential areas are mostly places with high-end residences, as well people that are either high or medium income settlers. Therefore, Break and Enter crime offenders target such places since there is the possibility of finding something valuable in such neighbourhoods. Moreover, the population aged 25 to 44 have an increasing effect on Break and Enter crimes, and that could be linked to low income for such age group. Unemployment information could not be obtained to give further information about the age group.

However, an increase in residents aged 45 to 64 reduces Break and Enter crimes. The reduction can be attributed to the economic stability of most of people within that age group as they rarely engage in Break and Enter crimes.

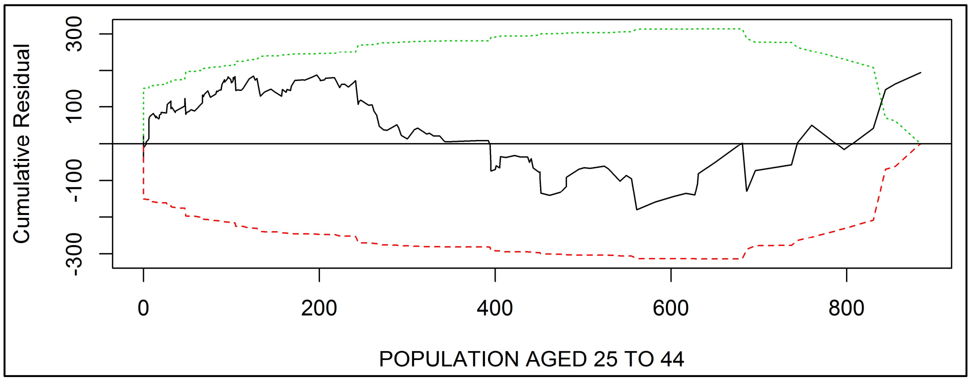


Figure 5.13: Cumulative Residual Plot for Model 3: Break and Enter Crimes

Break and Enter crimes were predicted by using estimates from the selected model. Empirical Bayes was then applied to the predicted values to estimate the expected numbers of Break and Enter crimes. Figure 5.14 is a Traffic Analysis Zone hotspot map for expected numbers of Break and Enter crimes, and Traffic Analysis Zones are ranked as high, medium, and low based on the expected numbers of Break and Enter crimes.

Table 5.33: Break and Enter Crimes Negative Binomial Regression Model Results

Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-3.39E-02	1.28E-01	-0.2652	0.7909	Dispersion Parameter = 1.7954
log1p(COMMERCIAL_AREA)	3.89E-02	1.42E-02	2.7350	0.0062	Standard Error = 0.207
POPULATION_DENSITY	9.92E-05	4.41E-05	2.2492	0.0245	Log-likelihood = -1455.106
LOW_DENSITY_RESIDENTIAL_AREA	2.28E-06	5.96E-07	3.8206	<0.001	
RETAIL_SPACE	2.22E-05	4.60E-06	4.8241	<0.001	
POPULATION_25TO44	5.77E-03	9.65E-04	5.9864	<0.001	
POPULATION_45TO64	-4.98E-03	1.06E-03	-4.6835	<0.001	

Mathematical representation of the selected Break and Enter crimes model is shown in Equation 5.7. Equations representing the top six Break and Enter crimes models are presented in Appendix H.

$$BREAK\_AND\_ENTER\_CRIMES = N \times \exp(-0.0339) \times \exp\begin{pmatrix} (\log COMMERCIAL\_AREA \times 0.0389) + (POPULATION\_DENSITY \times 9.92 \times 10^{-5}) + \\ (LOW\_DENSITY\_RESIDENTIAL\_AREA \times 2.28 \times 10^{-6}) + (RETAIL\_SPACE \times 2.22 \times 10^{-5}) + \\ (POPULATION\_25TO44 \times 5.77 \times 10^{-3}) + (POPULATION\_45TO64 \times -4.98 \times 10^{-3}) \end{pmatrix} \quad 5.7$$

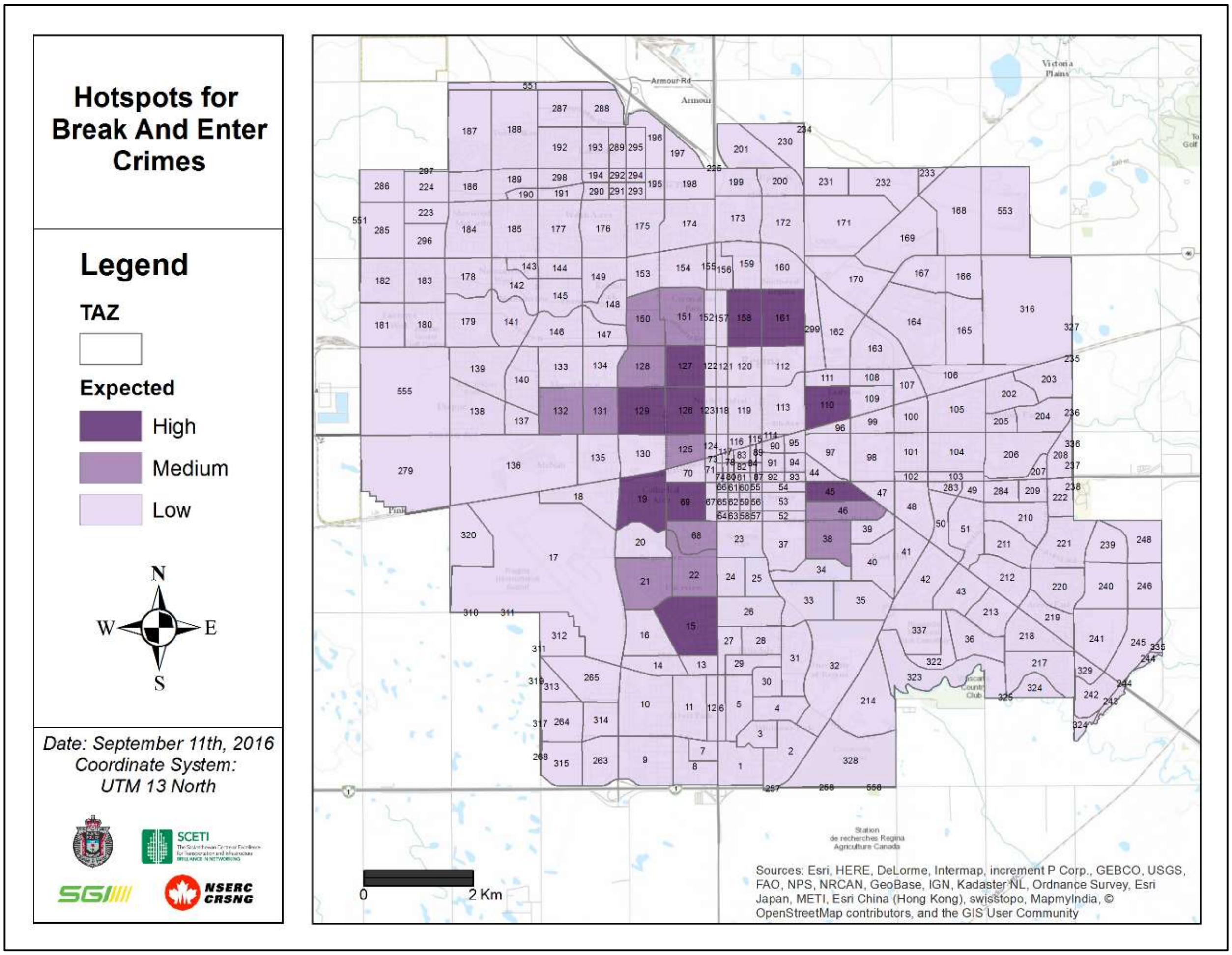


Figure 5.14: Hotspot Map for Expected Number of Break and Enter Crimes

Table 5.34 gives the range of values of expected numbers of Break and Enter crimes that constituted the three levels of hotspots as labelled in the map in Figure 5.14. Table 5.35 also provides further information about the top 10 high-level hotspots for expected numbers of Break and Enter crimes.

Table 5.34: Expected Numbers of Break and Enter Crimes per Traffic Analysis Zone Legend

Map Legend	Number of Break and Enter Crimes	Number of Traffic Analysis Zones
High	92.28 – 284.46	10
Medium	61.56 - 84.10	10
Low	Less than 60	242

Table 5.35: Top 10 Break and Enter Crimes Hotspots

Rank	Traffic Analysis Zone Number	Number of Expected Break and Enter Crimes
1	126	289.47
2	129	204.47
3	127	168.16
4	69	144.46
5	161	120.50
6	158	105.01
7	110	100.46
8	15	98.26
9	45	93.97
10	19	92.28

5.2.5 Mischief Crimes

Several models were created to predict mischief crimes, and the best 6 were chosen based on the results of the goodness-of-fit tests that were applied to them. Table 5.36 is a summary of the goodness-of-fit tests for the 6 candidate models. Model 1 provided the best predictive performance for mischief and was therefore the selected final mischief crime model. Table 5.37 provides the information about the sets of predictor variables used in each of the top six Mischief crimes models. P-values corresponding to these predictor variables are also presented in Table 5.37. As evident in Table 5.37, some variables had p-values greater than 0.05 but were maintained because they improved the predictive capability of models. Most variables had p-values lea than 0.05.

Table 5.36: Summary of Result of goodness-of-fit Tests for Mischief Crimes

Model	AIC	BIC	Dispersion Parameter	MSE	MSPE	R2FT		MPB	MAD	RMSE
						Calibration Data	Validation Data			
1	**1659.67**	**1685.69**	**1.82**	**2071.91**	**898.75**	**0.55**	**0.62**	**2.90**	**19.25**	**45.52**
2	1676.98	1699.74	1.61	4599.48	1072.42	0.33	0.54	5.14	19.00	67.82
3	1722.19	1741.71	1.24	1935.00	1281.31	0.52	0.46	-0.27	22.71	43.99
4	1678.48	1704.50	1.63	3384.74	6647.72	0.43	0.05	11.08	26.26	58.18
5	1661.77	1687.78	1.81	2481.38	939.60	0.54	0.60	0.38	19.35	49.81
6	1667.68	1690.44	1.73	3045.28	1096.42	0.48	0.63	1.23	19.17	55.18

Table 5.37: Top 10 Mischief Crimes Models Covariates and their P-Values

Covariate	P-Value
Model 1	
(Intercept)	0.8352
log1p(COMMERCIAL_AREA)	<0.001
POPULATION_DENSITY	<0.001
LOW_DENSITY_RESIDENTIAL_AREA	<0.001
RETAIL_SPACE	<0.001
POPULATION_25TO44	<0.001
POPULATION_45TO64	0.0087
Model 2	
(Intercept)	0.1730
Residential_Proportion	0.1122
log1p(COMMERCIAL_AREA)	<0.001
INDUSTRY_SPACE	0.0053
OFFICE_SPACE	0.0775
POPULATION_25TO44	<0.001
Model 3	
(Intercept)	0.0909
LAND_USE_PER_TAZ	<0.001
POPULATION_25TO44	<0.001
POPULATION_DENSITY	<0.001
POPULATION_45TO64	0.0324
Model 4	
(Intercept)	0.3497
log1p(COMMERCIAL_AREA)	<0.001
INDUSTRY_SPACE	<0.001
OFFICE_SPACE	0.0567
POPULATION_25TO44	<0.001
POPULATION_65_PLUS	0.7422
Log_Population_Density	0.1590
Model 5	
(Intercept)	0.3462
log1p(COMMERCIAL_AREA)	<0.001
POPULATION_DENSITY	0.0028
POPULATION_25TO44	0.0021
POPULATION_45TO64	0.0082
POPULATION_18TO24	0.0025
RETAIL_SPACE	<0.001

Table 5.38: Top 10 Mischief Crimes Models Covariates and their P-Values [Cont'd]

Covariate	P-Value
Model 6	
(Intercept)	0.0709
log1p(COMMERCIAL_AREA)	<0.001
POPULATION_18TO24	0.0060
POPULATION_25TO44	<0.001
POPULATION_45TO64	<0.001
RETAIL_SPACE	<0.001

The Cumulative Residual plot for the selected Mischief crimes model is shown in Figure 5.15, which depicts a good predictive model within the +2 and -2 standard deviations over most of the range of the commercial area. The Cumulative Residual plot goes out of range after the 10 mark on the x-axis but showed a flat line almost along the zero line for the better part of the 0 to 6 mark. This display depicts a good predictive capability of the model. Cumulative Residual plots for the top six mischief crimes models are presented in Appendix D.

Summary of the selected mischief crime Negative Binomial regression results are shown in Table 5.38. Log-transformed commercial area, retail space, population density, low density residential area, and population of residents aged 25 to 44 years are positively associated with mischief crimes. Commercial areas and retail spaces increase mischief crimes and can be attributed to their high trip attractions as well as presence of public properties. Neighbourhoods with higher population density have more amenities and properties, such as parks, playgrounds, and shopping malls, which could explain why increasing population density has an increasing effect on mischief crimes. Moreover, low density residential areas are often neighbourhoods with residents that are either medium or high class income earners, attracting mischief crime offenders. This finding could explain the increasing effect low density residential areas have on mischief crimes. As previously explained, the active population age group of 25 to 44 years are among the low-income earners, which could be the reason they engage in mischief crimes, making them one of the strongest predictors of mischief crimes.

As expected, the residents aged 45 to 64 are predominantly economically stable and have established careers. Consequently, they do not engage in many criminal activities, leading the age group to have a negative effect on mischief crimes.

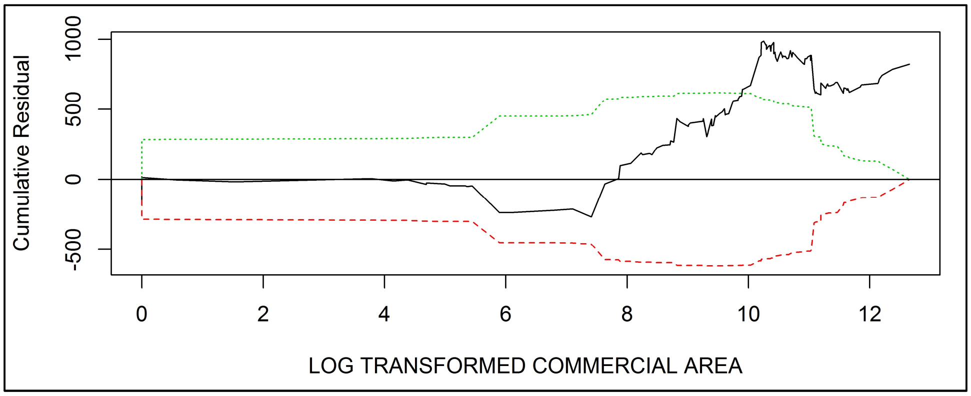


Figure 5.15: Cumulative Residual Plot for Selected Mischief Crimes

Table 5.38: Mischief Crimes Negative Binomial Regression Model Results

Covariate	Estimate	Standard Error	Z value	Pr(>|z|)	Model Statistics
(Intercept)	-2.62E-02	1.26E-01	-0.2081	0.8352	Dispersion Parameter = 1.8199 Standard Error = 0.215 Log-likelihood = -1643.669
log1p(COMMERCIAL_AREA)	7.17E-02	1.39E-02	5.1572	<0.001	
POPULATION_DENSITY	1.86E-04	4.29E-05	4.3297	<0.001	
LOW_DENSITY_RESIDENTIAL_AREA	2.19E-06	5.84E-07	3.7548	<0.001	
RETAIL_SPACE	2.21E-05	4.52E-06	4.8913	<0.001	
POPULATION_25TO44	4.12E-03	9.41E-04	4.3822	<0.001	
POPULATION_45TO64	-2.71E-03	1.03E-03	-2.6217	0.0087	

Equation 5.8 is a mathematical representation of the selected Mischief crimes model. Mathematical representation of the top six Mischief crimes models are presented in Appendix H.

$$MISCHIEF\_CRIMES = N \times \exp(-0.0262) \times \exp\begin{pmatrix} (\log COMMERCIAL\_AREA \times 0.0717) + (POPULATION\_DENSITY \times 1.86 \times 10^{-4}) + \\ (LOW\_DENSITY\_RESIDENTIAL\_AREA \times 2.19 \times 10^{-6}) + (RETAIL\_SPACE \times 2.21 \times 10^{-5}) + \\ (POPULATION\_25TO44 \times 4.12 \times 10^{-3}) + (POPULATION\_45TO64 \times -2.71 \times 10^{-3}) \end{pmatrix} \quad 5.8$$

Mischief crimes were then predicted using estimates from the selected Mischief crime model. Empirical Bayes method was then applied to the predicted numbers of mischief crimes to estimate expected numbers of mischief crimes. Figure 5.16 is a hotspot map for mischief crimes. The expected number of mischief crimes were ranked: the top 10 ranked Traffic Analysis Zones were defined as high; the top 11 to 20 Traffic Analysis Zones were defined as medium; and Traffic Analysis Zones ranked 21 and below were defined as low. Table 5.39 provides statistics for mischief crime hotspot Traffic Analysis Zones as shown in the map legend. Details of the expected number of mischief crimes for the top 10 hotspots labelled in the map as high are presented in Table 5.40.

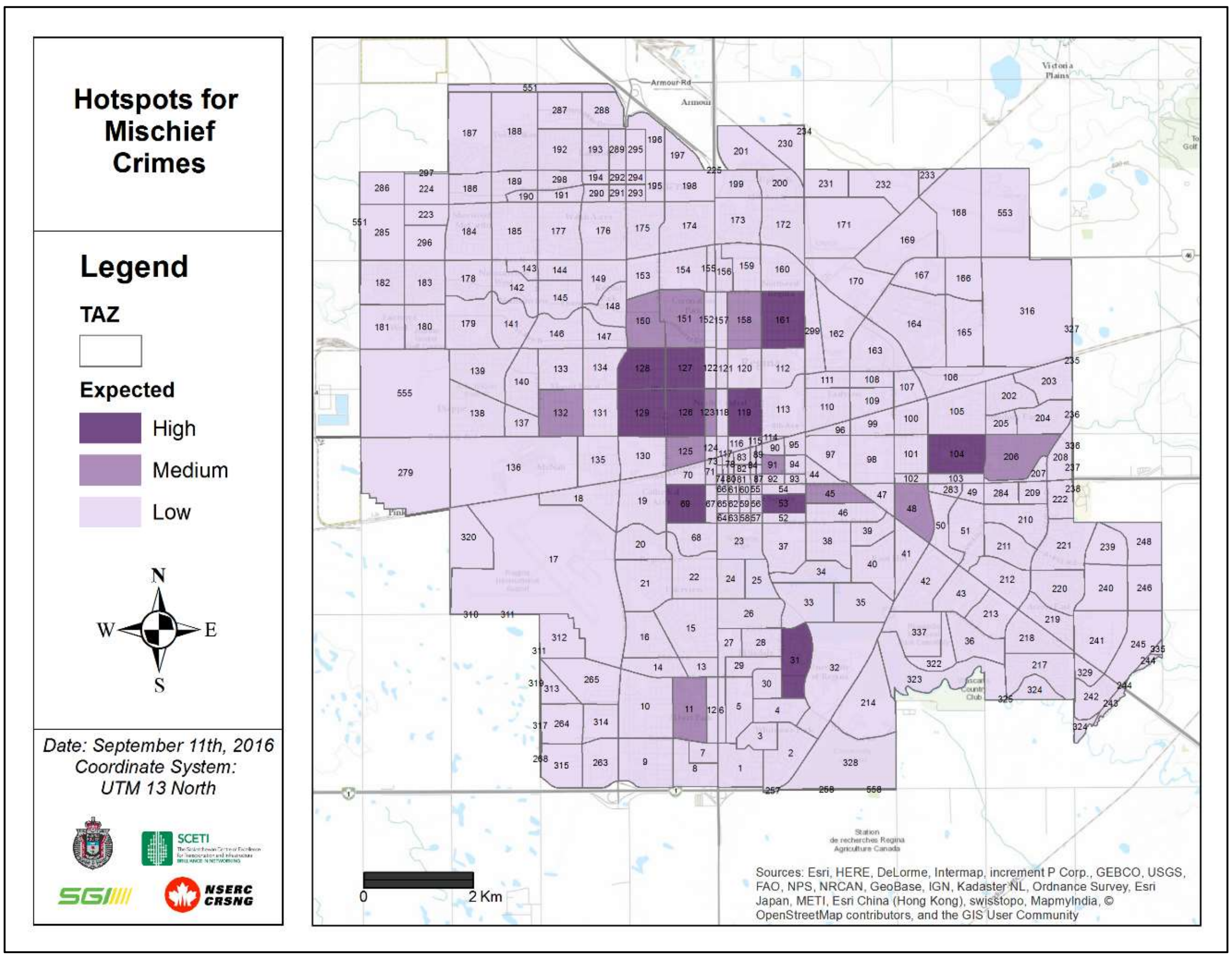


Figure 5.16: Hotspot Map for Expected Number of Mischief Crimes

Table 5.39: Expected number of Mischief Crimes per Traffic Analysis Zone Legend

Map Legend	Number of Mischief Crimes	Number of Traffic Analysis Zones
High	139.21 – 528.41	10
Medium	104.62 - 137.30	10
Low	Less than 98.06	242

Table 5.40: Top 10 Mischief Crimes Hotspots

Rank	Traffic Analysis Zone Number	Number of Expected Mischief Crimes
1	126	528.41
2	129	353.51
3	127	318.13
4	119	194.68
5	128	189.35
6	31	179.56
7	69	169.37
8	161	151.39
9	104	139.80
10	53	139.21

5.2.6 Theft Crimes

Several models were created to predict theft crimes, and the best 6 models were chosen as candidate models for the final theft model. The selection was based on results from the various goodness-of-fit tests that were applied to the models. Among the 6 candidate models, model number 1 had an overall best predictive performance and was selected as the final model. Table 5.41 summarizes the results from the goodness-of-fit tests results for the top six theft crimes models. The top six theft crimes models had different sets of predictor variables and Table 5.42 provides list of variables in each model. Most of these variables had p-values less than 0.05. Variables with had p-values greater than 0.05 were maintained because they improved significantly the predictive capability of models. Residents of population within the age group 25 to 44 years was the most significant predictor and was present in all models. Other variables that were significant in theft crime prediction included population age groups 18 to 24 years, 45 to 64 years and 65 years plus and were present in some models.

Table 5.41: Summary of Result of goodness-of-fit Tests for Theft Crimes

Model	AIC	BIC	Dispersion Parameter	MSE	MSPE	R2FT		MPB	MAD	RMSE
						Calibration Data	Validation Data			
1	**1673.70**	**1699.72**	**1.07**	**3897.57**	**15802.48**	**0.41**	**-0.36**	**12.59**	**46.86**	**62.43**
2	1626.51	1652.53	1.39	35887.28	6591.70	0.11	0.30	-1.64	31.35	189.44
3	1708.41	1734.42	0.90	115506.33	14643.73	-0.80	-0.08	8.15	42.87	339.86
4	1649.26	1675.28	1.23	183321.95	7466.44	-0.77	0.21	-0.66	34.10	428.16
5	1647.52	1673.54	1.24	221957.39	7419.35	-0.92	0.22	-0.79	33.48	471.12
6	1648.10	1670.86	1.23	233890.32	7020.77	-0.98	0.25	-1.20	32.25	483.62

Table 5.42: Top 10 Theft Crimes Models Covariates and their P-Values

Covariate	P-Value
Model 1	
(Intercept)	0.1532
log1p(COMMERCIAL_AREA)	<0.001
INDUSTRY_SPACE	0.0017
OFFICE_SPACE	<0.001
POPULATION_25TO44	<0.001
POPULATION_65_PLUS	0.1137
Log_Population_Density	0.3517
Model 2	
(Intercept)	0.0015
Residential_Proportion	<0.001
log1p(COMMERCIAL_AREA)	<0.001
POPULATION_18TO24	0.0548
POPULATION_25TO44	<0.001
POPULATION_45TO64	0.0027
RETAIL_SPACE	<0.001
Model 3	
(Intercept)	0.0058
LAND_USE_PER_TAZ	0.6991
POPULATION_25TO44	<0.001
POPULATION_DENSITY	0.0345
POPULATION_45TO64	0.0109
COMMERCIAL_AREA	<0.001
LOW_DENSITY_RESIDENTIAL_AREA	0.7664
Model 4	
(Intercept)	0.4554
log1p(COMMERCIAL_AREA)	<0.001
POPULATION_DENSITY	0.1229
LOW_DENSITY_RESIDENTIAL_AREA	0.7406
RETAIL_SPACE	<0.001
POPULATION_25TO44	0.0004
POPULATION_45TO64	0.0892
Model 5	
(Intercept)	0.3489
log1p(COMMERCIAL_AREA)	<0.001
POPULATION_DENSITY	0.0641
POPULATION_25TO44	0.0174
POPULATION_45TO64	0.0199
POPULATION_18TO24	0.1818
RETAIL_SPACE	<0.001

Table 5.42 Top 10 Theft Crimes Models Covariates and their P-Values [Cont'd]

Covariate	P-Value
Model 6	
(Intercept)	0.6544
log1p(COMMERCIAL_AREA)	<0.001
POPULATION_18TO24	0.2283
POPULATION_25TO44	<0.001
POPULATION_45TO64	0.0179
RETAIL_SPACE	<0.001

Cumulative Residual plot for the selected theft crimes model is shown in Figure 5.17, demonstrating good prediction close to the zero line and within the +2 and -2 standard deviations for the majority of the range of commercial area. Appendix D presents the Cumulative Residual plots for the top six theft crimes models. Results of the final selected theft crime Negative Binomial regression model are shown in Table 5.43.

Evident in Table 5.43, increase in log-transformed commercial area, industry area, and office area causes an increase in theft crimes. This increase can be due to the fact that such areas are places where many valuables can be found; therefore, an increase in such places in a Traffic Analysis Zone leads to an increase in expectations for theft crimes. Also, higher population density implies, higher numbers of people and properties, which is also expected to increase theft crimes. Again, population of residents aged 25 to 44 years may have lower median individual income or may have higher unemployment, influencing them to engage in theft to support themselves. Therefore, a higher number of such an age group will influence a higher number of theft crimes in a Traffic Analysis Zone.

However, the population of residents aged 65 and above, comprised mostly of retired citizens who are not physically active and are expected to enjoy their working life investments, do not engage in theft crimes. Accordingly, a higher population of residents aged 65 and over results in lower expected theft crimes in a Traffic Analysis Zone.

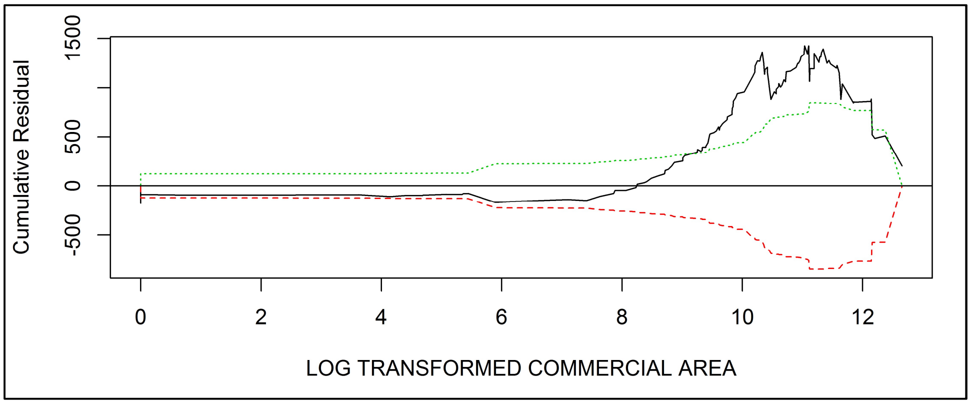


Figure 5.17: Cumulative Residual Plot Slected Thfet Crimes Model

Table 5.43: Theft Crimes Negative Binomial Regression Model Results

Covariate	Estimate	Standard Error	Z value	Pr(>|z|)	Model Statistics
(Intercept)	-4.63E-01	3.24E-01	-1.4284	0.1532	Dispersion Parameter = 1.0747
log1p(COMMERCIAL_AREA)	1.98E-01	1.69E-02	11.7228	<0.001	Standard Error = 0.113
INDUSTRY_SPACE	1.94E-05	6.18E-06	3.1347	0.0017	Log-likelihood = -1657.703
OFFICE_SPACE	2.35E-05	6.65E-06	3.5385	<0.001	
POPULATION_25TO44	2.33E-03	4.98E-04	4.6794	<0.001	
POPULATION_65_PLUS	-1.30E-03	8.25E-04	-1.5818	0.1137	
Log_Population_Density	4.73E-02	5.08E-02	0.9313	0.3517	

Equation 5.9 is a mathematical representation of the selected theft crimes model. Appendix H presents the mathematical representation of the top six theft crimes models.

$$THEFT\_CRIMES = N \times \exp(-0.463) \times \exp\begin{pmatrix} (\log COMMERCIAL\_AREA \times 0.198) + (INDUSTRY\_SPACE \times 1.94 \times 10^{-5}) + \\ (OFFICE\_SPACE \times 2.35 \times 10^{-5}) + (POPULATION\_25TO44 \times 2.33 \times 10^{-3}) + \\ (POPULATION\_65\_PLUS \times -1.3 \times 10^{-3}) + (\log POPULATION\_DENSITY \times 0.0473) \end{pmatrix} \quad 5.9$$

The variable estimates from the selected theft crimes model were then used to predict theft crimes; Empirical Bayes method was applied, and expected numbers of theft crimes were estimated. Figure 5.18 is a theft crime hotspot map by Traffic Analysis Zone for the City of Regina. Traffic Analysis Zones were ranked using the expected numbers of theft crimes, and the rankings were grouped into high, medium, and low, as illustrated in the map. The range of values of expected numbers of theft crimes for the three hotspot levels are presented in Table 5.44. Details about the 10 high hotspots for theft crimes are listed in Table 5.45 with their Traffic Analysis Zone numbers and expected numbers of theft crimes.

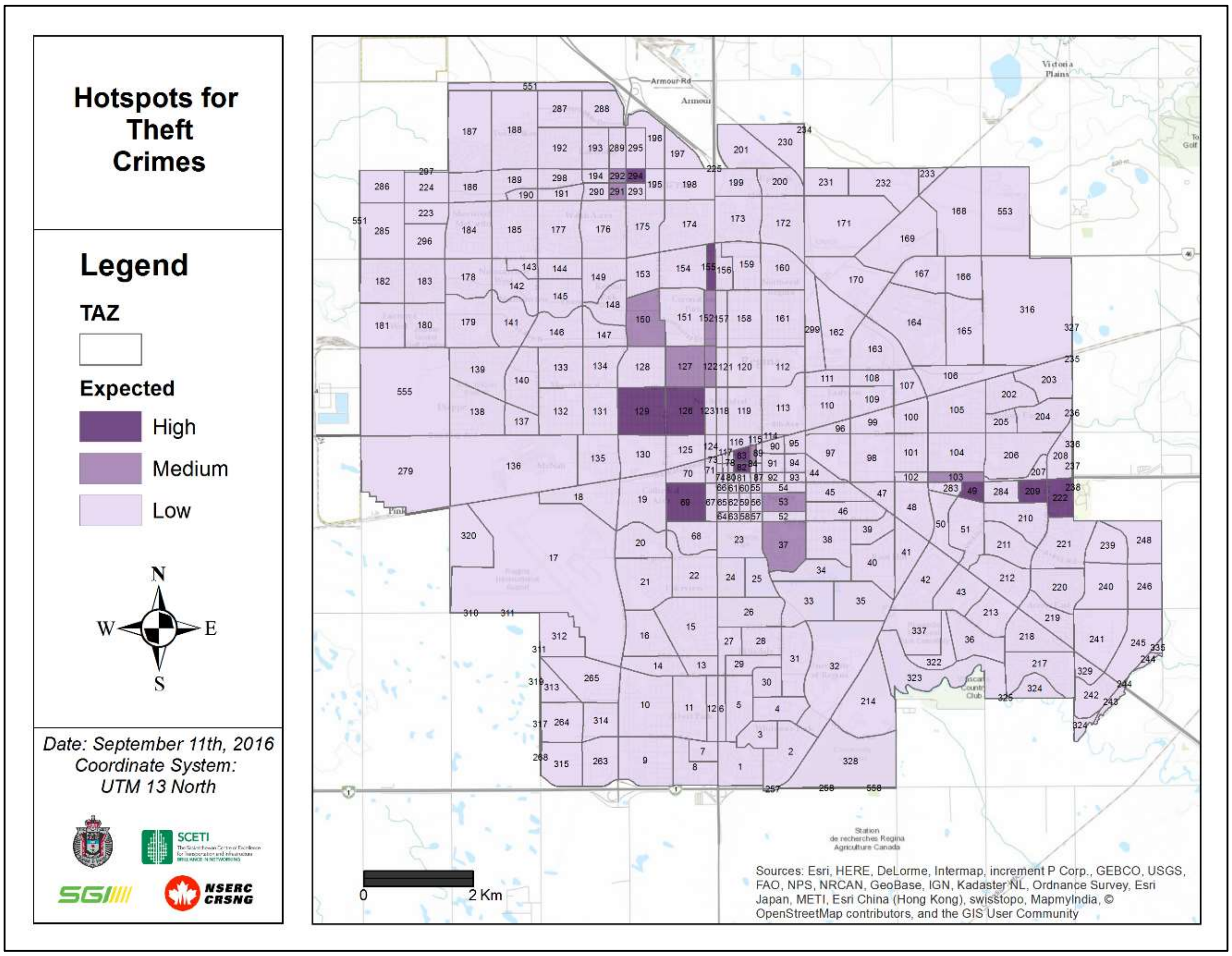


Figure 5.18: Hotspot Map for Expected Numbers of Theft Crimes

Table 5.44: Expected number of Theft Crimes per Traffic Analysis Zone Legend

Map Legend	Number of Theft Crimes	Number of Traffic Analysis Zones
High	203.98 – 585.09	10
Medium	136.71 - 202.14	10
Low	Less than 136.63	242

Table 5.45: Top 10 Theft Crimes Hotspots

Rank	Traffic Analysis Zone Number	Number of Expected Theft Crimes
1	83	585.09
2	155	581.11
3	209	380.29
4	82	379.11
5	222	369.85
6	126	254.45
7	294	253.41
8	49	240.79
9	129	233.70
10	69	203.98

5.2.7 Theft from Auto

Among the several models developed to predict Theft from Auto crimes, the best six were selected using the results from their goodness-of-fit test results. Table 5.46 presents the results of the goodness-of-fit test for these models; evidently, model 1 had the best predictive performance and was therefore, the selected final Theft from Auto crimes model. Table 5.47 presents the predictor variables and their corresponding p-values for the top six Theft from Auto crimes models. Most of the predictor variables the top six models had p-values below 0.05 but were maintained because of their effect on the predictive performance of models. Population of residents aged 25 to 44 years old were the most significant Theft from Auto crimes predictor and was present in all six models. In comparison to the other models, the top six Theft from Auto models predicted poorly.

Table 5.46: Summary of Result of goodness-of-fit Tests for Theft from Auto Crimes

Model	AIC	BIC	Dispersion Parameter	MSE	MSPE	R2FT		MPB	MAD	RMSE
						Calibration Data	Validation Data			
1	**1586.22**	**1615.49**	**1.61**	**598.87**	**20934.35**	**0.54**	**-1.12**	**16.67**	**33.94**	**24.47**
2	1634.03	1656.80	1.17	1008.55	645.31	0.37	0.43	1.05	18.41	31.76
3	1586.21	1612.23	1.59	691.53	545.63	0.51	0.62	0.18	14.75	26.30
4	1600.97	1626.99	1.45	898.52	27331.76	0.47	-1.32	20.31	34.06	29.98
5	1600.47	1626.48	1.46	808.23	472.81	0.49	0.61	-1.95	14.84	28.43
6	1599.63	1622.39	1.45	870.93	463.14	0.47	0.62	-1.88	14.47	29.51

Table 5.47: Top 10 Theft from Auto Crimes Models Covariates and their P-Values

Covariate	P-Value
Model 1	
(Intercept)	0.0004
POPULATION_25TO44	<0.001
INDUSTRY_SPACE	<0.001
POPULATION_DENSITY	<0.001
RETAIL_SPACE_PROP	<0.001
log1p(LOW_DENSITY_RESIDENTIAL_AREA)	0.0023
LAND_USE_PER_TAZ	<0.001
COMMERCIAL_AREA	0.1702
Model 2	
(Intercept)	<0.001
Residential_Proportion	<0.001
OFFICE_SPACE	0.7946
POPULATION_25TO44	<0.001
POPULATION_65_PLUS	0.0205
Log_Population_Density	0.7751
Model 3	
(Intercept)	0.7559
log1p(COMMERCIAL_AREA)	<0.001
POPULATION_DENSITY	0.0037
LOW_DENSITY_RESIDENTIAL_AREA	<0.001
RETAIL_SPACE	<0.001
POPULATION_25TO44	0.0071
POPULATION_45TO64	0.0732
Model 4	
(Intercept)	0.6829
log1p(COMMERCIAL_AREA)	<0.001
INDUSTRY_SPACE	<0.001
OFFICE_SPACE	0.3957
POPULATION_25TO44	<0.001
POPULATION_65_PLUS	0.1053
Log_Population_Density	0.4316
Model 5	
(Intercept)	0.1727
log1p(COMMERCIAL_AREA)	<0.001
POPULATION_DENSITY	0.2746
POPULATION_25TO44	0.0248
POPULATION_45TO64	0.3279
POPULATION_18TO24	0.0962
RETAIL_SPACE	<0.001

Table 5.47 Top 10 Theft from Auto Crimes Models Covariates and their P-Values [Cont'd]

Covariate	P-Value
Model 6	
(Intercept)	0.0819
log1p(COMMERCIAL_AREA)	<0.001
POPULATION_18TO24	0.1071
POPULATION_25TO44	0.0116
POPULATION_45TO64	0.3006
RETAIL_SPACE	<0.001

Figure 5.19 is the Cumulative Residual plot for the selected Theft from Auto crimes model and it shows a model with a good prediction within the +2 and -2 standard deviations across the range of population aged 25 to 44. Appendix D presented the Cumulative Residual plots for the top six Theft from Auto crimes models.

The selected Theft from Auto crimes model regression results are presented in Table 5.48. All predictors were positively associated with theft from auto crimes. Industry spaces, retail spaces, and commercial areas are locations with large parking spaces and, therefore, attract theft from auto crimes, thus explaining their positive effect on theft from auto crimes. Similarly, Traffic Analysis Zones with multiple or mixed land uses would have a significant numbers of parking lots, and, as such, Traffic Analysis Zones with higher numbers of mixed land use are expected to have higher theft from auto crimes. Also, places with higher population density means more parking lots for the residents, and, intuitively, that increases theft from auto crimes. As mentioned previously, low density residential areas have medium and high class income dwellers, and such places are targeted for high-valued items in their vehicles, increasing theft from auto crimes. Moreover, the population age group 25 to 44 years are among the lowest income earners. Therefore, to supplement their earnings or support themselves, residents in that age group may engage in Theft from Auto crimes.

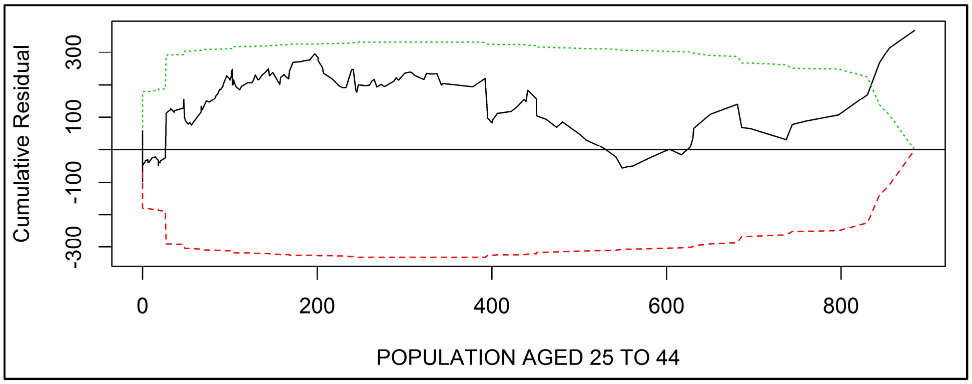


Figure 5.19: Cumulative Residual Plot for the Selected Theft from Auto Crimes Model

Table 5.48: Theft from Auto Crimes Negative Binomial Regression Model Results

Covariate	Estimate	Standard Error	Z value	Pr(>|z|)	Model Statistics
(Intercept)	-7.26E-01	2.04E-01	-3.5624	0.0004	Dispersion Parameter = 1.6087 Standard Error = 0.195 Log-likelihood = -1568.219
POPULATION_25TO44	1.97E-03	3.90E-04	5.0400	<0.001	
INDUSTRY_SPACE	1.96E-05	4.20E-06	4.6684	<0.001	
POPULATION_DENSITY	1.30E-04	4.20E-05	3.0865	<0.001	
RETAIL_SPACE_PROP	4.19E+00	7.72E-01	5.4293	<0.001	
log1p(LOW_DENSITY_RESIDENTIAL_AREA)	5.72E-02	1.88E-02	3.0473	0.0023	
LAND_USE_PER_TAZ	1.78E-01	5.09E-02	3.5007	<0.001	
COMMERCIAL_AREA	1.90E-06	1.38E-06	1.3717	0.1702	

Equation 5.10 is a mathematical representation of the selected Theft from Auto crimes model. Appendix H presents mathematical representation of the top six Theft from Auto crimes models.

$$THEFT\_FROM\_AUTO\_CRIMES = N \times \exp(-0.726) \times \exp\begin{pmatrix} (POPULATION\_25TO44 \times 1.97 \times 10^{-3}) + (INDUSTRY\_SPACE \times 1.96 \times 10^{-5}) + \\ (POPULATION\_DENSITY \times 1.3 \times 10^{-4}) + (RETAIL\_SPACE\_PROP \times 4.19) + \\ (\log LOW\_DENSITY\_RESIDENTIAL\_AREA \times 0.0572) + (LAND\_USE\_PER\_TAZ \times 0.178) + \\ (COMMERCIAL\_AREA \times 1.9 \times 10^{-6}) \end{pmatrix} \quad 5.10$$

Numbers of Theft from Auto crimes were then predicted using estimates from the selected theft from auto crimes model.

Expected numbers of Theft from Auto crimes were determined by applying the Empirical Bayes method to the predicted numbers. Figure 5.20 is a theft from auto crimes hotspot map, and Table 5.49 provides range of values of expected numbers of Theft from Auto crimes for the legend in the map in Figure 5.20. Traffic Analysis Zones labelled as high were the top ten ranked locations by the expected number of theft from auto crimes; Traffic Analysis Zones labelled as medium were the top 11 to 20 ranked locations; and all other locations were labelled as low. Table 5.50 provides statistics about the top 10 high theft from auto hotspot Traffic Analysis Zones.

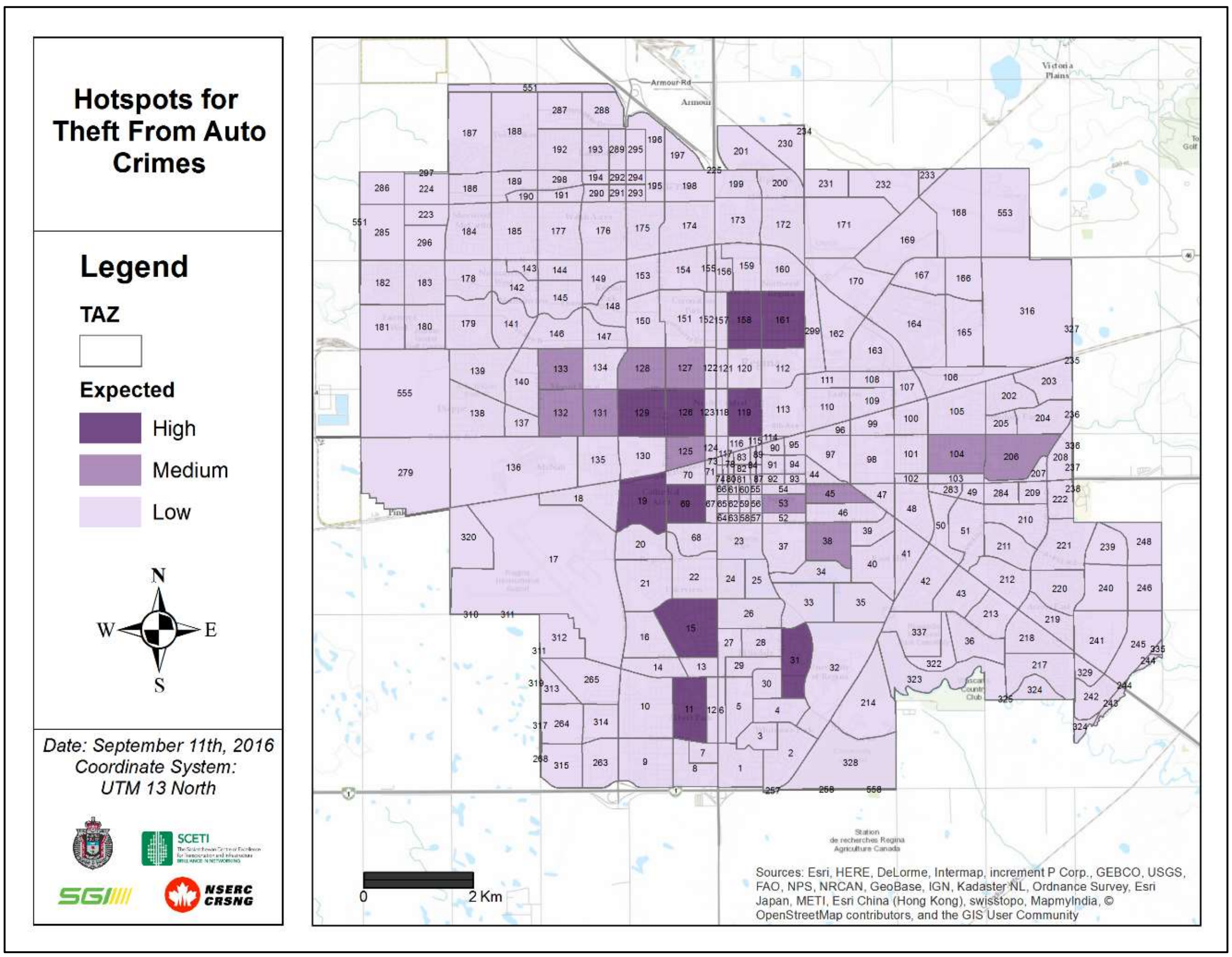


Figure 5.20: Hotspot Map for Expected Number of Theft from Auto Crimes

Table 5.49: Expected number of Theft from Auto Crimes per Traffic Analysis Zone Legend

Map Legend	Number of Theft from Auto Crimes	Number of Traffic Analysis Zones
High	96.69 – 174.47	10
Medium	80.30 - 92.81	10
Low	Less than 79.6	242

Table 5.50: Top 10 Theft from Auto Crimes Hotspots

Rank	Traffic Analysis Zone Number	Number of Expected Theft from Auto Crimes
1	119	174.47
2	126	161.53
3	129	152.41
4	158	121.95
5	161	121.20
6	31	114.41
7	15	109.18
8	69	105.29
9	19	105.08
10	11	96.69

5.2.8 Theft of Auto

Out of the six candidate models for theft of auto crimes, model 1 had the overall best predictive performance and was the selected Theft of Auto crimes models. Goodness-of-fit test results for the top six Theft of Auto crimes models are presented in Table 5.51. These six individual crimes had different sets of predictor variables and Table 5.52 presents the variables in these six models and their corresponding p-values. All the predictor variables in models 1 and 2 had p-values less than 0.05. However, in the remaining other models some variables had p-values greater than 0.05 but were maintained because they improved the predictive performance of models.

Table 5.51: Summary of Result of goodness-of-fit Tests for Theft of Auto Crimes

Model	AIC	BIC	Dispersion Parameter	MSE	MSPE	R2FT		MPB	MAD	RMSE
						Calibration Data	Validation Data			
1	**1350.35**	**1379.62**	**1.97**	**691.53**	**202.30**	**0.61**	**0.58**	**0.98**	**8.81**	**26.30**
2	1400.64	1420.15	1.33	1008.55	490.87	0.39	0.30	0.62	11.63	31.76
3	1387.45	1416.72	1.48	598.87	246909.26	0.42	-9.69	61.66	69.76	24.47
4	1375.95	1401.97	1.58	898.52	62851.19	0.39	-4.47	33.91	40.25	29.98
5	1379.14	1405.15	1.60	808.23	188.14	0.53	0.58	1.29	8.68	28.43
6	1378.75	1401.51	1.58	870.93	187.70	0.51	0.60	1.47	8.64	29.51

Table 5.52: Top 10 Theft of Auto Crimes Models Covariates and their P-Values

Covariate	P-Value
Model 1	
(Intercept)	<0.001
log1p(COMMERCIAL_AREA)	0.0337
POPULATION_DENSITY	0.0177
RETAIL_SPACE	<0.001
POPULATION_25TO44	<0.001
POPULATION_45TO64	<0.001
LAND_USE_PER_TAZ	<0.001
URBAN_HOLDING_RESIDENTIAL_AREA	0.0155
Model 2	
(Intercept)	<0.001
log1p(COMMERCIAL_AREA)	0.0028
POPULATION_DENSITY	<0.001
LAND_USE_PER_TAZ	<0.001
URBAN_HOLDING_RESIDENTIAL_AREA	0.0071
Model 3	
(Intercept)	<0.001
POPULATION_25TO44	<0.001
INDUSTRY_SPACE	<0.001
log1p(POPULATION_DENSITY)	0.0157
RETAIL_SPACE_PROP	<0.001
log1p(LOW_DENSITY_RESIDENTIAL_AREA)	0.4469
LAND_USE_PER_TAZ	<0.001
COMMERCIAL_AREA	0.0078
Model 4	
(Intercept)	0.0016
log1p(COMMERCIAL_AREA)	<0.001
INDUSTRY_SPACE	<0.001
OFFICE_SPACE	0.8617
POPULATION_25TO44	<0.001
POPULATION_65_PLUS	0.6372
Log_Population_Density	0.5903
Model 5	
(Intercept)	0.0030
log1p(COMMERCIAL_AREA)	<0.001
POPULATION_DENSITY	0.1956
POPULATION_25TO44	<0.001
POPULATION_45TO64	<0.001
POPULATION_18TO24	0.2483
RETAIL_SPACE	<0.001

Table 5.52 Covariates and P-values for Top 6 Theft of Auto Crimes Models [Cont'd]

Covariate	**P-Value**
Model 6	
(Intercept)	0.0061
log1p(COMMERCIAL_AREA)	<0.001
POPULATION_18TO24	0.3103
POPULATION_25TO44	<0.001
POPULATION_45TO64	<0.001
RETAIL_SPACE	<0.001

The Cumulative Residual plot for the selected Theft of Auto crime model is shown in Figure 5.21. However, the Cumulative Residual plot does not show the best fit to the data. Because the standard deviation values are quite small, this finding could explain why, comparatively, it has the best predictive capability for theft of auto crimes Cumulative Residual plots for all the top six Theft of Auto crimes models are presented in Appendix D. Table 5.53 presents results of the Negative Binomial regression for the selected theft of auto crimes model. From the regression results, log-transformed commercial area, population density, retail space, population of residents aged 25 to 44 years, and number of land user per Traffic Analysis Zone were positively associated with theft of auto crimes, implying that an increase in these variables results in an increase of theft of auto crimes. Intuitively, commercial areas, high population density areas, retails spaces, and TAZs with high numbers of mixed land use will have relatively higher numbers of parking spaces, implying there will be more vehicles. Therefore, it is no surprise that all those variables have a positive effect on theft of auto crimes.

On the other hand, increases in urban holding residential areas and populations of residents aged 45 to 64 result in reduction of theft of auto crimes. Population of residents aged 45 to 64 are mostly comprised of economically stable people, and, therefore, the likelihood of engaging in criminal activities decreases. Moreover, urban holding residential areas are lands that have not been developed for dwellings nor any forms of commercial activities. Therefore, such lands do not attract trips, resulting in almost no parking of vehicles, which explains why increases in urban holding residential area land use results in reduction of theft of auto crimes.

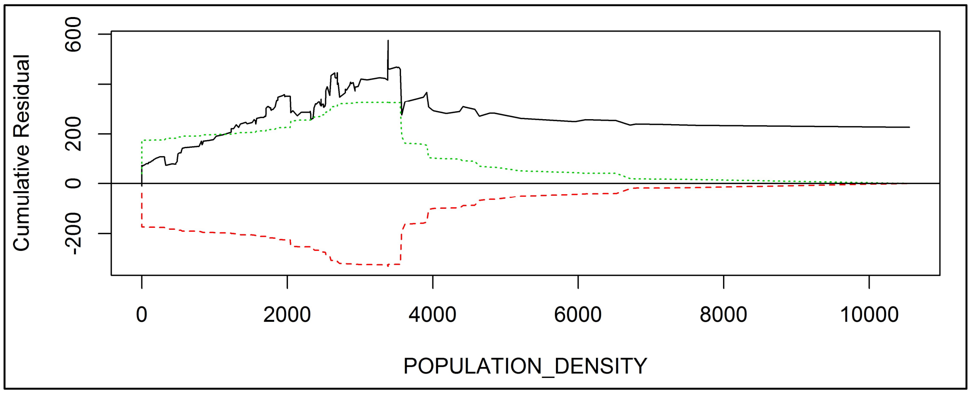


Figure 5.21: Cumulative Residual Plot for Model 4: Theft of Auto Crimes

Table 5.53: Theft of Auto Crimes Negative Binomial Regression Model Results

Covariate	Estimate	Standard Error	Z value	Pr(>|z|)	Model Statistics
(Intercept)	-1.06E+00	2.00E-01	-5.3010	<0.001	Dispersion Parameter = 1.9673
log1p(COMMERCIAL_AREA)	3.17E-02	1.49E-02	2.1231	0.0337	Standard Error = 0.261
POPULATION_DENSITY	9.88E-05	4.17E-05	2.3715	0.0177	Log-likelihood = -1332.349
RETAIL_SPACE	1.97E-05	4.48E-06	4.4017	<0.001	
POPULATION_25TO44	5.64E-03	9.48E-04	5.9511	<0.001	
POPULATION_45TO64	-3.88E-03	9.46E-04	-4.0996	<0.001	
LAND_USE_PER_TAZ	2.38E-01	4.30E-02	5.5292	<0.001	
URBAN_HOLDING_RESIDENTIAL_AREA	-6.09E-07	2.52E-07	-2.4214	0.0155	

Equation 5.11 provides a mathematical representation of the selected Theft of Auto crime model. Appendix H presents mathematical representations of the top six theft of auto crimes models.

$$THEFT\_OF\_AUTO\_CRIMES = N \times \exp(-1.06) \times \exp\begin{pmatrix} (\log COMMERCIAL\_AREA \times 0.0317) + (POPULATION\_DENSITY \times 9.88 \times 10^{-5}) + \\ (RETAIL\_SPACE \times 1.97 \times 10^{-5}) + (POPULATION\_25TO44 \times 5.64 \times 10^{-3}) + \\ (POPULATION\_45TO64 \times -3.88 \times 10^{-3}) + (LAND\_USE\_PER\_TAZ \times 0.238) + \\ (URBAN\_HOLDING\_RESIDENTIAL\_AREA \times 6.09 \times 10^{-7}) \end{pmatrix} \tag{5.11}$$

Numbers of theft of auto crimes were predicted using estimates from the selected theft of auto crimes model.

The expected numbers of theft of auto crimes were estimated using Empirical Bayes approach, and the hotspots map for the expected numbers of theft of auto crimes is shown in Figure 5.22. Table 5.54 provides the description for high, medium, and low in the hotspot map. Furthermore, Table 5.55 provides details about the ranked top 10 high hotspot Traffic Analysis Zones as well as the numbers of expected theft of auto crimes.

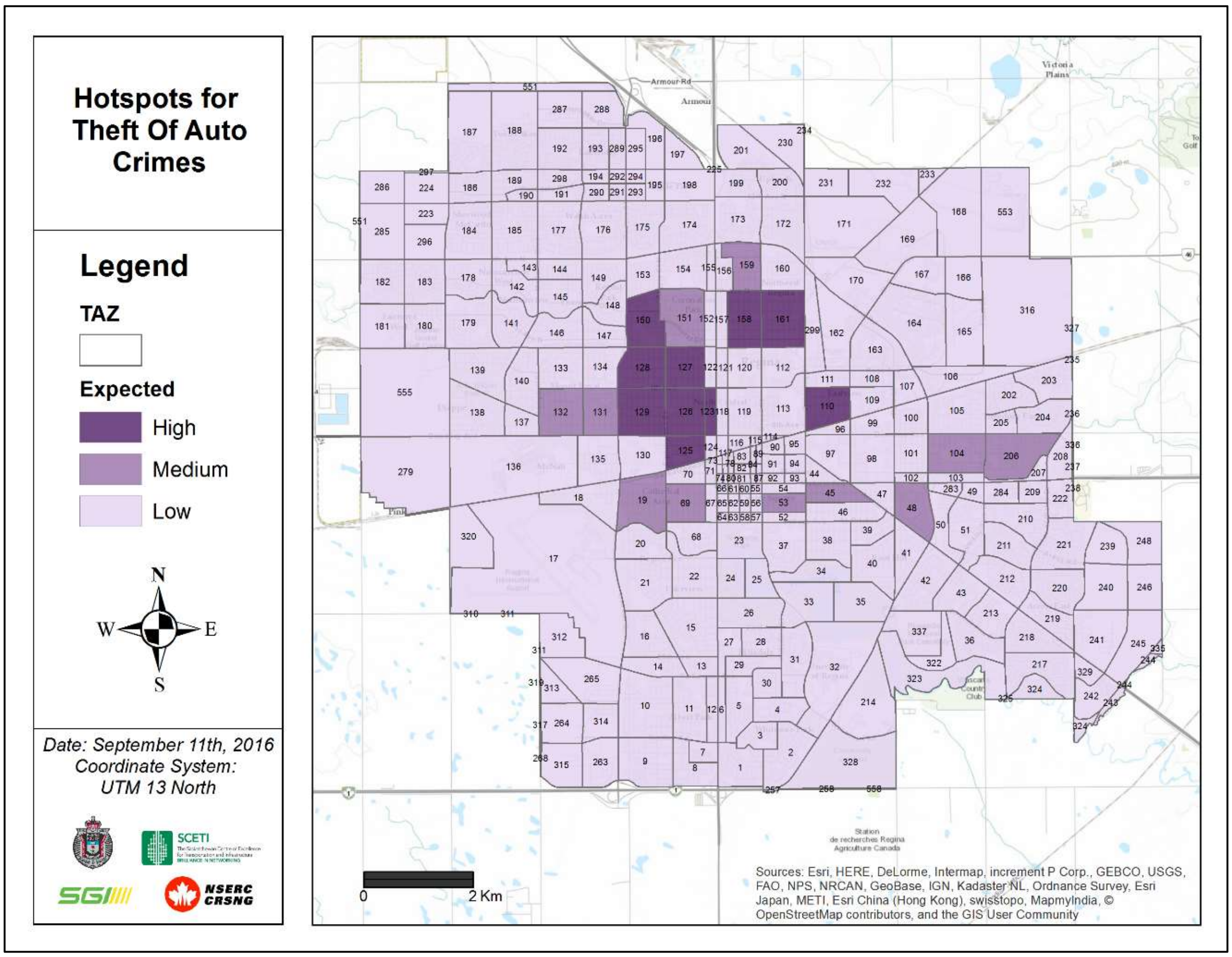


Figure 5.22: Hotspot Map for Expected Number of Theft of Auto Crimes

Table 5.54: Expected number of Theft of Auto Crimes per Traffic Analysis Zone Legend

Map Legend	Number of Theft of Auto Crimes	Number of Traffic Analysis Zones
High	67.43 – 238.29	10
Medium	47.56 - 67.24	10
Low	Less than 47.06	242

Table 5.55: Top 10 Theft of Auto Crimes Hotspots

Rank	Traffic Analysis Zone Number	Number of Expected Theft of Auto Crimes
1	126	238.21
2	129	175.53
3	127	172.60
4	128	94.74
5	158	85.08
6	161	80.46
7	125	76.43
8	123	73.49
9	150	69.38
10	110	67.43

5.3 Data-Driven Approaches to Crime and Traffic Safety (DDACTS) Zone Maps

As previously explained, after developing models for both traffic collisions and crimes, Traffic Analysis Zones with overlap of high occurrence of both collisions and crimes are of interest in this research. Therefore, a further exploratory analysis was done to determine these Traffic Analysis Zones with overlap of high numbers of expected collisions and crimes. These overlap areas will be referred to as DDACTS zones. Figure 5.23 is a map that depicts DDACTS Zone maps for Total collisions and violent crimes. All the DDACTS Zone maps are presented in Appendix F. However, Tables with the expected frequencies of collisions and crimes per DDACTS Zone are presented in this chapter. The DDACTS Zone maps are grouped into three, based on collision severities: total, fatal-injury, and property damage only.

5.3.1 Total Collisions and Different Crime Types DDACTS Zones

Figure 5.23 is a map representing the DDACTS zone for total collisions and violent crimes, and Table 5.56 provides details about the frequencies of expected total collisions and violent crimes.

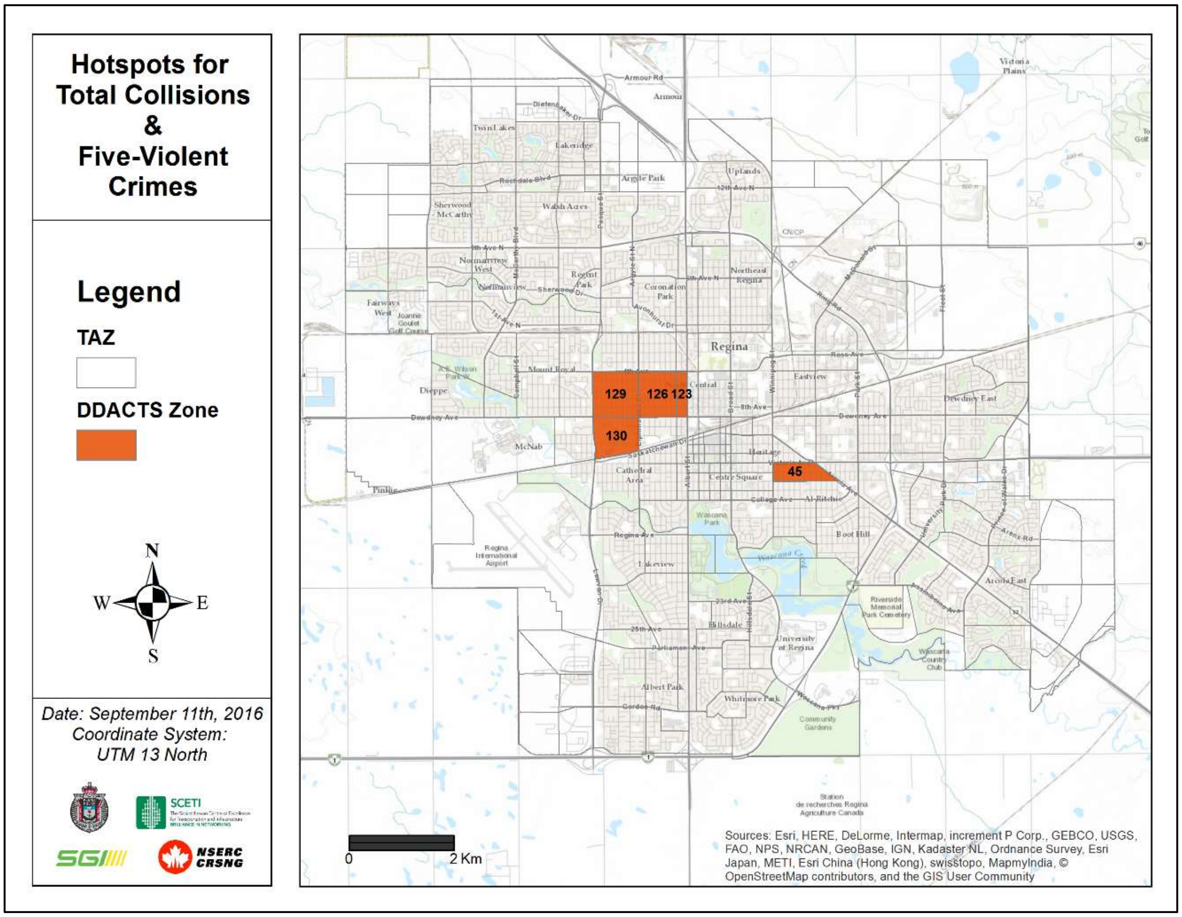


Figure 5.23: DDACTS Zone Map: Total Collision and Violent Crimes

Table 5.56: DDACTS Zone Statistics for Total Collisions and Violent Crimes

Traffic Analysis Zone Number	Expected Total Collisions	Expected Violent Crimes
126	424.27	942.30
129	403.32	480.03
123	366.36	311.26
130	345.91	154.89
45	284.81	136.99

Table 5.57 provides details about the frequencies of expected total collisions and assault crimes.

Table 5.57: DDACTS Zone Table for Total Collisions and Assault Crimes

Traffic Analysis Zone Number	Expected Total Collisions	Expected Assault Crimes
126	424.27	683.23
129	403.32	335.51
123	366.36	228.56
130	345.91	113.63

Table 5.58 provides details about the frequencies of expected total collisions and robbery crimes.

Table 5.58: DDACTS Zone Statistics for Total Collisions and Robbery Crimes

Traffic Analysis Zone Number	Expected Total Collisions	Expected Robbery Crimes
126	424.27	942.30
129	403.32	480.03
123	366.36	311.26
130	345.91	154.89
45	284.81	136.99

Table 5.59 provides details about the frequencies of expected total collisions and Break and Enter crimes.

Table 5.59: DDACTS Zone Statistics for Total Collisions and Break and Enter Crimes

Traffic Analysis Zone Number	Expected Total Collisions	Expected Break and Enter Crimes
126	424.27	289.47
129	403.32	204.47
19	362.14	92.28
45	284.81	93.97

Table 5.60 provides details about the frequencies of expected total collisions and mischief crimes.

Table 5.60: DDACTS Zone Statistics for Total Collisions and Mischief Crimes

Traffic Analysis Zone Number	Expected Total Collisions	Expected Mischief Crimes
126	424.27	528.41
129	403.32	353.51
119	351.81	194.68
104	311.24	139.80

Table 5.61 provides details about the frequencies of expected total collisions and theft crimes.

Table 5.61: DDACTS Zone Statistics for Total Collisions and Theft Crimes

Traffic Analysis Zone Number	Expected Total Collisions	Expected Theft Crimes
126	424.27	254.45
129	403.32	233.70

Table 5.62 provides details about the frequencies of expected total collisions and theft from auto crimes.

Table 5.62: DDACTS Zone Statistics for Total Collisions and Theft from Auto Crimes

Traffic Analysis Zone Number	Expected Total Collisions	Expected Theft from Auto Crimes
126	424.27	161.53
129	403.32	152.41
19	362.14	105.08
119	351.81	174.47

Table 5.63 provides details about the frequencies of expected total collisions and theft of auto crimes.

Table 5.63: DDACTS Zone Statistics Table for Total Collisions and Theft of Auto Crimes

Traffic Analysis Zone Number	Expected Total Collisions	Expected Theft of Auto Crimes
126	424.27	238.21
129	403.32	175.53
123	366.36	73.49

5.3.2 Fatal-Injury Collisions and Different Crime Types DDACTS Zones

Table of values of expected frequencies of fatal-injury collisions and crimes for DDACTS Zone maps are presented in this section. Maps depicting these zones are presented in appendix F.

Table 5.64 provides details about the frequencies of expected Fatal-Injury collisions and violent crimes.

Table 5.64: DDACTS Zone Statistics for Fatal-Injury Collisions and Violent Crimes

Traffic Analysis Zone Number	Expected Fatal-Injury Collisions	Expected Violent Crimes
126	103.07	942.30
130	89.58	154.89
123	80.52	311.26
129	77.68	480.03

Table 5.65 provides details about the frequencies of expected Fatal-Injury collisions and assault crimes.

Table 5.65: DDACTS Zone Table for Fatal-Injury Collisions and Assault Crimes

Traffic Analysis Zone Number	Expected Fatal-Injury Collisions	Expected Assault Crimes
123	80.52	228.56
126	103.07	683.23
129	77.68	335.51
130	89.58	113.63

Table 5.66 provides details about the frequencies of expected Fatal-Injury collisions and robbery crimes.

Table 5.66: DDACTS Zone Statistics for Fatal-Injury Collisions and Robbery Crimes

Traffic Analysis Zone Number	Expected Fatal-Injury Collisions	Expected Robbery Crimes
126	103.07	139.30
130	89.58	21.97
123	80.52	57.08
129	77.68	92.57

Table 5.67 provides details about the frequencies of expected Fatal-Injury collisions and Break and Enter crimes.

Table 5.67: DDACTS Zone Statistics for Fatal-Injury Collisions and Break and Enter Crimes

Traffic Analysis Zone Number	Expected Fatal-Injury Collisions	Expected Break and Enter Crimes
126	103.07	289.47
19	80.68	92.28
129	77.68	204.47

Table 5.68 provides details about the frequencies of expected Fatal-Injury collisions and mischief crimes.

Table 5.68: DDACTS Zone Statistics for Fatal-Injury Collisions and Mischief Crimes

Traffic Analysis Zone Number	Expected Fatal-Injury Collisions	Expected Mischief Crimes
126	103.07	528.41
129	77.68	353.51
104	72.80	139.80

Table 5.69 provides details about the frequencies of expected Fatal-Injury collisions and theft crimes.

Table 5.69: DDACTS Zone Statistics for Fatal-Injury Collisions and Theft Crimes

Traffic Analysis Zone Number	Expected Fatal-Injury Collisions	Expected Theft Crimes
126	103.07	254.45
129	77.68	233.70

Table 5.70 provides details about the frequencies of expected Fatal-Injury collisions and theft from auto crimes.

Table 5.70: DDACTS Zone Statistics for Fatal-Injury Collisions and Theft from Auto Crimes

Traffic Analysis Zone Number	Expected Fatal-Injury Collisions	Expected Theft from Auto Crimes
126	103.07	161.53
129	77.68	152.41

Table 5.71 provides details about the frequencies of expected Fatal-Injury collisions and theft of auto crimes.

Table 5.71: DDACTS Zone Statistics for Fatal-Injury Collisions and Theft of Auto Crimes

Traffic Analysis Zone Number	Expected Fatal-Injury Collisions	Expected Theft of Auto Crimes
126	103.07	238.21
123	80.52	73.49
129	77.68	175.53

5.3.3 Property Damage Only Collisions and Different Crime Types DDACTS Zones

Tables representing frequencies of expected Property Damage Only collisions and crimes for DDACT zone maps are presented in this section. Maps for these zones are presented in appendix F. Table 5.72 provides details about the frequencies of expected Property Damage Only collisions and violent crimes.

Table 5.72: DDACTS Zone Statistics for Property Damage Only Collisions and Violent Crimes

Traffic Analysis Zone Number	Expected Property Damage Only Collisions	Expected Violent Crimes
129	326.36	480.03
126	322.43	942.30
123	285.29	311.26
130	254.92	154.89
45	225.78	136.99

Table 5.73 provides details about the frequencies of expected Property Damage Only collisions and assault crimes.

Table 5.73: DDACTS Zone Statistics for Property Damage Only Collisions and Assault Crimes

Traffic Analysis Zone Number	Expected Property Damage Only Collisions	Expected Assault Crimes
129	326.36	335.51
126	322.43	683.23
123	285.29	228.56
130	254.92	113.63

Table 5.74 provides details about the frequencies of expected Property Damage Only collisions and robbery crimes.

Table 5.74: DDACTS Zone Statistics for Property Damage Only Collisions and Robbery Crimes

Traffic Analysis Zone Number	Expected Property Damage Only Collisions	Expected Robbery Crimes
129	326.36	92.57
126	322.43	139.30
123	285.29	57.08
130	254.92	21.97
45	225.78	22.46
69	214.75	23.21

Table 5.75 provides details about the frequencies of expected Property Damage Only collisions and Break and Enter crimes.

Table 5.75: DDACTS Zone Statistics for Property Damage Only Collisions and Break and Enter Crimes

Traffic Analysis Zone Number	Expected Property Damage Only Collisions	Expected Break and Enter Crimes
129	326.36	204.47
126	322.43	289.47
19	281.20	92.28
45	225.78	93.97
69	214.75	144.46

Table 5.76 provides details about the frequencies of expected Property Damage Only collisions and mischief crimes.

Table 5.76: DDACTS Zone Statistics for Property Damage Only Collisions and Mischief Crimes

Traffic Analysis Zone Number	Expected Property Damage Only Collisions	Expected Mischief Crimes
129	326.36	353.51
126	322.43	528.41
119	308.27	194.68
104	236.26	139.80
69	214.75	169.37

Table 5.77 provides details about the frequencies of expected Property Damage Only collisions and theft crimes.

Table 5.77: DDACTS Zone Statistics for Property Damage Only Collisions and Theft Crimes

Traffic Analysis Zone Number	Expected Property Damage Only Collisions	Expected Theft Crimes
129	326.36	233.70
126	322.43	254.45
69	214.75	203.98

Table 5.78 provides details about the frequencies of expected Property Damage Only collisions and theft from auto crimes.

Table 5.78: DDACTS Zone Statistics for Property Damage Only Collisions and Theft from Auto Crimes

Traffic Analysis Zone Number	Expected Property Damage Only Collisions	Expected Theft from Auto Crimes
129	326.36	152.41
126	322.43	161.53
119	308.27	174.47
19	281.20	105.08
69	214.75	105.29

Table 5.68 provides details about the frequencies of expected Property Damage Only collisions and theft of auto crimes.

Table 5.79: DDACTS Zone Statistics for Property Damage Only Collisions and Theft of Auto Crimes

Traffic Analysis Zone Number	Expected Property Damage Only Collisions	Expected Theft of Auto Crimes
129	326.36	175.53
126	322.43	238.21
123	285.29	73.49

5.4 Chapter Summary

Results of goodness-of-fit tests of the top selected models were presented in this chapter. There were top 10 collision models and top 6 crime models. Based on the results of the goodness-of-fit tests, the overall best model was selected as the final model for the three collision severities: total, fatal-injury, and property damage only collisions. The overall best predictive crime model based on the goodness-of-fit test results were also determined for the different crime occurrence types: violent, assault, robbery, break and enter, mischief, theft, theft from auto, and theft of auto crimes. Regression results for the final models were also presented in this chapter, and explanations were provided for the outcome of these models. Hotspot maps representing the expected numbers of collisions and crimes were created by using results from Empirical Bayes technique. Statistics for these hotspots and for the top 10 Traffic Analysis Zone hotspots were also presented. An exploratory analysis was performed to identify Traffic Analysis Zones that have high occurrences of both collisions and crimes. These Traffic Analysis Zones were then called DDACTS zones. Maps were created for all DDACTS zones, and the statistics for these zones presented.

CHAPTER 6 . CONCLUSION

This research was focussed on developing a data-driven approach by employing DDACTS, which is a fairly new, advanced, and innovative approach to simultaneously deal with the issue of traffic collisions and crimes on a Traffic Analysis Zone-level (macro-level). The unit of analysis for this research was Traffic Analysis Zones. By integrating multiple databases into Traffic Analysis Zone level, several models were developed to predict collision and crime per Traffic Analysis Zone. Typically, a macro-level analysis is done to provide valuable safety considerations to be made in assigning land use, road network characteristics, socioeconomic, and demographics to a much larger unit of analysis such as neighbourhoods, Traffic Analysis Zone, wards, county, region, country *etc.*. However, in this research, a further step was taken by employing the Empirical Bayes method, which combined information from both the predicted and observed collision and crime frequencies to estimate expected numbers of incidents at a macro-level. The estimated expected frequencies of collisions and crimes can then be used for enforcement and implementation of safety countermeasures. It is worthy of mention that the conclusions drawn from this research are based on the data split percentages for calibration and validation. For collision prediction, 90% of total data were used in model calibration and 10% were used in data validation. 70% of total crime data were used for calibration and the remaining 30% were used in model validation.

6.1 Collision Prediction Models for the City of Regina

In total, there were 30 final candidate models developed. For each severity type, 10 models were determined, and the best model was selected: Total, Fatal-injury, and Property Damage Only collisions. Final selected models provided important information about Traffic Analysis Zone level road safety at both the planning stage and in determining areas that require countermeasures. Conclusions drawn from this research include the following:

- Both intersection density and intersection road density had positive associations with collisions. Higher numbers of these variables resulted in higher collision frequencies; as such, they provide information about some safety concerns. Intersections, therefore, should be provided only when necessary.
- When comparing 3-leg and 4-leg intersections, 3-leg intersections had fewer safety concerns.

- Road corridors with posted speed limits of 80km/hr provided few safety concerns, potentially due to the fact that high speed roadways have less congestion.
- Low density residential areas have collision reduction effects.

The developed models can be used as tools in the neighbourhood planning stages in various ways:

- Using socio-demographic trends, such as population, population density, and residential density, measures of safety can be determined for Traffic Analysis Zones and necessary interventions can be implemented.
- Measures of safety can be estimated using road network information.
- Land use types employed in models can help planners generate scenarios to determine and improve neighbourhood road safety.
- To help safety initiatives, planners can generate multiple scenarios with different variables, thus, allowing them to evaluate the safety effects of each scenario.

6.2 Crime Prediction Models for the City of Regina

The created crime models provided information about how land use type, socio-demographics, and residential land use type influence different crime types. Some conclusion drawn include the following:

- Commercial areas and retail spaces were targeted areas for a high numbers of violent crimes.
- High population density neighbourhoods attracted high numbers of crimes.
- Higher numbers of residents within the age groups of 18 to 24 and 25 to 44 were positively associated with both violent and non-violent crimes.
- Residents within the age groups of 45 to 65 as well as 65 years and over had a crime reduction effect, regardless of the crime occurrence type.
- Low density residential areas attracted many non-violent crimes
- Industry and office areas also attracted many non-violent crimes.
- Multiple or mixed land use areas also attracted a high numbers of auto-involving theft crimes.

6.3 Data-Driven Approaches to Crime and Traffic Safety (DDACTS) Zones for the City of Regina

Some of the conclusions drawn from the DDACTS zone maps are:

- Employing the DDACTS zone maps alongside clockplots showing peak hours of collisions and crimes provides a powerful tool for conducting focussed enforcement targeted as specific incidents.
- The downtown area (North Central Neighbourhood) was a high hotspot for different crime types, especially in Traffic Analysis Zones 123, 126, 119, and 129 along Dewdney Avenue. Therefore, those areas can be considered for the appropriate countermeasures.
- Some Traffic Analysis Zones along Saskatchewan drive, including 130, 19, and 69, experienced high numbers of both collisions and non-violent crimes.
- Traffic Analysis Zone 45, particularly along Arcola drive, should be a target area for both traffic and criminal intervention.
- Property Damage Only collisions and robbery crimes had the highest numbers of DDACTS zone hotspots, implying robbery crimes occur in close proximity to occurrences of Property Damage Only collisions.

6.4 Future Work and Recommendations

This research has determined some proactive approaches to simultaneously reducing collisions and crimes with the aim of reducing social harm. Some recommendation for future work to advance this research include the following points:

- Collecting and employing further Traffic Analysis Zone level data, including unemployment by age group and demographic data by gender, in model development to determine the effect of such data on the predictive performance of developed models.
- Combination of qualitative and quantitative crime prediction in the future. Each approach compliments the other, and the joint approach would be a powerful tool in crime prediction models.
- Obtaining Traffic Analysis Zone level information by time would provide further information about the temporal pattern of collision and crime.

- Weather and road conditions data can be incorporated in models to determine the influence they will have on collisions and crime prediction models.

In conclusion, this research presents important data-driven road and neighbourhood safety considerations to be made at the planning stage of new neighbourhoods, as well as the focus areas to be considered for enforcement to reduce both collisions and crimes. These two inputs can save lives both on the road and in neighbourhoods and can improve the quality of lives of residents.

REFERENCES

Abdel-Aty, M., Siddiqui, C., & Huang, H. (2011). Managing roadway safety at the traffic analysis level. *International Conference on Transportation Information and Safety, 2011*. Wuhan, China: American Society of Civil Engineers.

Aguero-Valverde, J., & Jovanis, P. P. (2006). Spatial analysis of fatal and injury crashes in Pennsylvania. *Accident Analysis and Prevention, 1262*, 1-9.

Bates, L., Watson, & King, M. (2006). Driver education and graduated driver licensing. *Proceedings Australasian Road Safety, Policing and Education Conference,* Gold Coast.

Boyce, J., Cotter, A., & Perreault, S. (2014). Police-reported crime statistics in Canada, 2013. *Statistics Canada catalogue no. 85-002-X, ISSN 1209-6393.*

Brännäs, K., & Johansson, P. (1992). Time series count data regression. *Umea Economic Studies,* 289.

Browning, R, C. Byron, A. C., Calder, A. C., Krivo, J. L., Kwan, M. Lee, J., & Peterson, D. R. (2010). Commercial density, residential concentration, and crime: Land use patterns and violence in neighborhood context. *Journal of Research in Crime and Delinquency, 47*(3), 329-357.

Burrow, L., & Taylor, M. (1995). Network safety management—A quantitative approach. *Proceedings of the 23rd European Transport Forum, Seminar G,* 115-126. Wokingham, UK: PTRC Education and Research Services Limited.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? — Arguments against avoiding RMSE in the literature. *Geoscientific Model Development, 7*, 1247-1250.

Chang, L. (2005). Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network. *Safety Science, 43*, 541-557.

Chen, L., Chen, C., & Ewing, R. (2012). The relative effectiveness of pedestrian safety countermeasures at urban intersections — Lessons from a New York City experience. *Transportation Research Board 91st Annual Meeting*. Washington, DC: Transportation Research Board.

City of Regina (n.d). OpenData Source Website: http://open.regina.ca/organization/city-of-regina

City of Regina. (2011). National Household Survey. Retrieved from http://www.regina.ca/opencms/export/sites/regina.ca/residents/city-planning/.media/pdf/2015-neighbourhood-profiles/downtown.pdf

City of Thunder Bay. (2010). Creating Cycling Education Programs: The Tookit.

Cohen, J., & Gorr, W. L. (2005). Development of crime forecasting and mapping systems for use by police (Final Report) Carnegie Mellon University, Pittsburgh, PA.

Criminal Code, R.S.C. 1985, c.46.

Datta, G. S., Kubokawa, T., Molina, I., & Rao, J. N. K. (2010). Estimation of mean squared error of model-based small area estimators. *Test, 20*(2), 368-388.

Estochen, M, B., Strauss, T., & Souleyrette, R, D. (1998). An assessment of emergency response vehicle pre-deployment using GIS identification of high-accident density locations. *1998 Transportation Conference Proceedings.* Ames, Iowa: Center for Transportation Research and Education.

Fell, J. C. (2013). The effects of increased traffic enforcement on other crimes. *2013 Australian Road Safety Research, Policing & Education Conference.* Calverton, MD: Pacific Institute for Research and Evaluation.

Fridstrøm, L., Ifver, J., Ingebrigtsen, S., Kulmala, R., & Thomsen, L. K. (1995). Measuring the contribution of randomness, exposure, weather, daylight to the variation in road accidents counts. *Accident Analysis & Prevention, 27*(1), 1-20.

Fridstrøm, L., & Ingebrigtsen, S. (1991). An aggregate accident model based on pooled, regional time-series data. *Accident Analysis & Prevention, 23*(5), 363-378.

Giacopassi, D., & Forde, D. R. (2000). Broken windows, crumpled fenders, and crime. *Journal of Criminal Justice 28(5),* 397-405.

Glenn, S. (2012). Coefficient of Determination (R Squared): Definition and Calculation. Retrieved from http://www.statisticshowto.com/what-is-a-coefficient-of-determination/

Glenn, S. (2012). Spearman Rank Correlation (Spearman's Rho): Definition and How to Calculate it. Retrieved from http://www.statisticshowto.com/spearman-rank-correlation-definition-calculate/

Government of Saskatchewan. (2013). Selective traffic enforcement program (STEP). Retrieved from http://highways.gov.sk.ca/step/

Grabowski, D. C., & Morrisey, M. A. (2007). System wide implications of the repeal of the national maximum speed limit. *Accident Analysis & Prevention, 39*, 180-189.

Hadayeghi, A. (2009). Use of advanced techniques to estimate zonal level safety planning models and examine their temporal transferability (Doctoral dissertation). University of Toronto, Toronto.

Hadayeghi, A., Shalaby, A. S., & Persuad, B. (2003). Macrolevel crash prediction models for evaluating safety of urban transportation systems. *Transportation Research Record, 1897,* 87-95.

Hadayeghi, A., Shalaby, A. S., Persuad, B. (2007). Safety prediction models: Proactive tool for safety evaluation in urban transportation planning applications. *Transportation Research Record, 2019*, 225-236.

Hadayeghi, A., Shalaby, A. S., Persuad, B., & Cheung, C. (2006). Temporal transferability and updating of zonal level accident prediction models. *Accident Analysis & Prevention, 38*, 579-589.

Hakim, S., Shefer, D., Hakkert, A. S., & Hockerman, I. (1991). A critical review of macro models for road accidents. *Accidents Analysis & Prevention, 23*(5), 379-400.

Hamidi, A., Fontaine, M. D., & Demetksy, M. J. (2010). A planning-level methodology for identifying high-crash sections of Virginia's primary system (Research report). Charlottesville, VA: Virginia Transportation Research Council.

Hauer, E. (1992) Empirical Bayes approach to the estimation of "unsafety": The multivariate regression Mmethod, *Accident Analysis & Prevention*, *24*(5), 457-477.

Hauer, E. (1997) *Observational before-after studies in road safety: Estimating the effect of highway and traffic engineering measures on road safety*. Oxford, UK: Elservier.

Hauer, E. (2001). Overdispersion in modelling accidents on road sections and in Empirical Bayes estimation. *Accident Analysis & Prevention, 33*, 799-808.

Hauer, E. (2004). The harm done by tests of significance. *Accident Analysis & Prevention, 36*, 495-500.

Hauer, E. (2015). The art of regression modeling in road safety. Cham, Switzerland: Springer.

Hauer, E & Bamfo, J. (1997) Two tools for finding what function links the dependent variable to the explanatory variables. *Proceedings of ICTCT 97 Conference*, Lund, Sweden.

Hauer, E, Harwood D. W., and Council, F. (2002) Estimating safety by the Empirical Bayes method: A tutorial. *Transportation Research Record, 1784*, 126-131. Washington, D.C.: Transportation Research Board.

Heim, K. (2014). Visualization and modeling for crime data indexed by road segments (Doctoral dissertation. George Mason University, Fairfax, VA.

Heywood, D. I., Cornelius, S., & Carver, S. (1998). *An introduction to geographical systems.* New York: Addison Wesley Longman.

Huang, H., Song, B., Xu, P., Zeng, Q., Lee, J., & Abdel-Aty, M. (2016). Macro and micro models for zonal crash prediction with application in hot zones identification. *Journal of Transport Geography, 54*, 248-256.

Ihssian, A. (2014). Investigation of the influence of boundary data assignment on the development of multimodal macro-level Collision Prediction Models (CPMs) (Masters thesis). Carleton University, Ottawa, ON.

Jovanis, P. P., & Chang, H. (1986). Modeling the relationship of accidents to miles traveled. *Transportation Research Record, 1068*, 42-51.

Kaplan, S., & Prato, C. G. (2007). Impact of BAC limit reduction on different population segments: A poisson fixed effect analysis. *Accident Analysis & Prevention, 34,* 357-365.

Kim, Y, T., Kim, H, K., & Park, H, B. (2010). Accident models of trumpet interchange s-type ramps using by Poisson, negative binomial regression and ZAM. *KSCE Journal of Civil Engineering, 15*(3), 545-551.

Kmet, L., & Macarthur, C. (2006). Urban-rural differences in motor vehicle crash fatality and hospitalization rates among children and youth. *Accident Analysis & Prevention, 38*(1)*,* 122-127.

Kopits, E., & Cropper, M. (2005). Traffic fatalities and economic growth. *Accident Analysis & Prevention, 37,* 169-178.

Kumara, S.S., & Chin, H. C. (2004). Study of fatal traffic crashes in Asia Pacific countries. *Transportation Research Record: Journal of the Transportation Research Board, 1897,* 43-47.

Leigh, J. P., & Waldon, H. M. (1993). Unemployment and highway fatalities. *Journal of Health Politics: Policy and Law, 16*(1), 135-155.

Liu, X., Saat, R, M., Qin, X., & Barkan, P, L, C. (2013). Analysis of U.S freight-train derailment severity using zero-truncated negative binomial regression and quantile regression. *Accident Analysis & Prevention, 59,* 87-93.

Laffont, J, L., Wright, D., Alarcon, C., Grothaus, D., Remund, K., & Gregoire, S. (2013). Introduction to Heteroscedastic Linear Model and Generalized Linear Model Application: Testing for Biotech Traits. https://www.seedtest.org/upload/cms/user/ISTAGLM.pdf

Leader Post, (Bohmann Rikkeal). (2014).

http://www.leaderpost.com/life/Regina+population+cracks/9553267/story.html

Lochner, L., & Moretti, E. (2003). The effects of education on crimes: Evidence from prison inmates, arrests, and self-reports. *American Economic Review, 94*(1), 155-189.

Lovegrove, G. R. (2005) Community-based, macro-level collision prediction models (Doctoral Dissertation). University of British Columbia, Canada.

Lovegrove, G., Lim, C., & Sayed, T. (2010) Community-based, macrolevel collision prediction model use with a regional transportation plan. *Journal of Transportation Engineering, 136*(2), 120-128.

Lovegrove, G. R., & Litman, T. (2007) Using macro-level collision prediction modes to evaluate the road safety effects of mobility management strategies: New empirical tools to promote sustainable development (CD-ROM). *2008 Transportation Research Board Annual Meeting.* Washington, DC.

Lovegrove, G. R., and Sayed, T. (2005) Macro-level collision prediction models for evaluating neighbourhood traffic safety. *6th Transportation Specialty Conference*. Toronto, ON: Canadian Society of Civil Engineering.

Lovegrove, G. R., and Sayed, T. (2006) Using macro-level collision prediction models in road safety planning applications. *Transportation Research Record: Journal of Transportation Research Board, 1950,* 73-82.

Machin, S., Marie, O., & Vujic, S. (2010). The crime reducing effect of education (Discussion paper). Bonn, Germany: Institute for the Study of Labour (IZA).

Majumdar, A., Noland, R. B., & Ochieng, W.Y. (2004). A spatial and temporal analysis of safety-belt usage by simulation. *Accident Analysis & Prevention, 36,* 1162-1169.

Mayhew, D. R., & Simpson, H. M. (2002). The safety value of driver education and training. *Injury Prevention 2002, 8*, ii3-ii8.
Michalowski, R. J. (1975). Violence in the road: the crime of vehicular homicide. *Journal of Research in Crime and Delinquency, 12(1),* 30-43

McGuigan, D. R. D. (1982) Non-junction collision rates and their use in 'Black-Spot' identification, *Traffic Engineering & Control*, *23*, 60-65.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology 47,* 90-100.

Naderan, A., & Shahi, J. (2010). Aggregate crash prediction models: Introducing crash generation concept. *Accident Analysis & Prevention, 42*, 339-346.

National Highway Traffic Safety Administration. (2014). Data-driven approaches to crime and traffic safety (DDACTS): Operational guidelines revised. Meridian, ID: IADLEST.

National Highway Traffic Safety Administration. (1997). The Peoria experience, traffic enforcement and crime: It plays in Peoria. Retrieved from http://www.nhtsa.gov/people/injury/enforce/peoria/peoria.pdf

Nau, R. (2014). Notes on linear regression analysis. Retrieved from http://people.duke.edu/~rnau/Notes_on_linear_regression_analysis--Robert_Nau.pdf

Nicholson, A.J. (1980) Identification of hazardous locations. *PRU Newsletter, 66*. Auckland, New Zealand: Ministry of Transport.

Noland, R. B. (2003). Medical treatment and traffic fatality reductions in industrialized countries. *Accident Analysis & Prevention, 35,* 877-883.

Noland, R. B., & Oh, L. (2004). The effect of infrastructure and demographic change on traffic-related fatalities and crashes: A case study of Illinois county-level data. *Accident Analysis & Prevention, 36,* 525-532.

Noland, R. B., & Quddus, M. A. (2004). Improvements in medical care and technology and reductions in traffic-related fatalities in Great Britain. *Accident Analysis & Prevention, 36,* 973-984.

Osgood, D., W. (2000). Poisson-based regression analysis of aggregate crime rates. *Journal of Quantitative Criminology, 16*(1), 21-43.

Persaud, B., Lord, D., & Palmisano, J. (2002) Calibration and transferability of accident prediction models for urban intersections. *Transportation Research Record, 1784*, 57-64. Washington, DC: TRB National Research Council.

Perry, L. W., McInnis, B., Price, C. C., Smith, C. S., & Hollywood, S. J. (2013). Predictive policing: The role of crime forecasting in law enforcement operations. Santa Monica, CA: RAND.

Peters, A., & MacDonald, H. (2004). *Unblocking the Census with GIS,* ESRI Publication, 2004.

Pratt, C. Travis. (2001). Assessing the relative effects of macro-level predictors of crime: A meta-analysis (Doctoral dissertation). University of Cincinnati.

Pulugurtha, S, Srinivas., Duddu, R. Venakata., & Kotagiri, Yashaswi. (2013). Traffic analysis zones level crash estimation models based on land use characteristics. *Accident Analysis & Prevention, 50*, 678-687.

Retting, A. R., Ferguson, A. S., & McCartt, A. T. (2003). A review of evidence-based traffic engineering measures to reduce pedestrian-motor vehicle crashes (CD-ROM). *2004 Transportation Research Board Annual Meeting.* Chicago, IL.

Royal Canadian Mounted Police (RCMP). (2014). RCMP in Saskatchewan year in review 2013-2014. Retrieved from http://www.rcmp-grc.gc.ca/sk/publications/yearinreview-retrospective delannee-2013-14/safe-roads-eng.htm

Saskatchewan Government Insurance (SGI). (2013). *Traffic safety in Saskatchewan* [PDF document]. Regina, SK: SGI.

Saskatchewan Government Insurance (SGI). (2011). *Report Impaired Drivers Program* [News Release Retrived from https://www.sgi.sk.ca/about/responsibility/trafficsafety/rid.html]. Regina, SK: SGI.

Sawalha, Z. A. (2002). Traffic Accident Modeling: Statistical Issues and Safety Applications (Doctoral dissertation). University of British Columbia.

Schneider, S. (2002). Predicting crime: A review of the research (Summary Report). Ryerson University.

Schnelle, J. F., Kirchner, R. E., Casey, J. D., Uselton, P. H., & McNees, M. P. (1977). Patrol evaluation research: a multiple-baseline analysis of saturation police patrolling during day and night hours. *Journal of Applied Behavior Analysis, 10,* 33-40.

Shen, Q. (2007). Development of safety performance functions for Empirical Bayes estimation of crash reduction factors (Doctoral dissertation). Florida International University.

Shimizu, H., & Desrochers, P. (2015). *Speed or Greed: Does automated traffic enforcement improve safety or generate revenue?* [PDF document]. Retrieved from https://fcpp.org/sites/default/files/documents/Shimizu%20%26%20Desrochers%20-%20Speed%20or%20Greed.pdf

Shingleton, S., J. (2012). Crime trend prediction using regression models for Salinas, California (Masters thesis). Naval Postgraduate School.

Smeed, R. J. (1949). Some statistical aspects of road safety research. *Journal of the Royal Statistical Society: Series A (General), 112* (1), 1-34.

Stark, R. (1987). Deviant Places: a theory of the ecology of crime. *Criminology, 25(4),* 893-910.

Statistics Canada Demography Division. (2013). Annual demographic estimates: Canada, provinces and territories. Ottawa, ON: Minister of Industry.

Street, J, T., Winter, D., Buckley, S., Nicholson, P., & Twomey, A. (1999). Trauma on rural roads: the role of a peripheral hospitality. *Injury: international Journal of the Care of the Injured, 20,* 337-340.

Stucky, D, T., & Ottensmann, R., J. (2009). Land use and violent crime. *American Society of Criminology,* 47, 1223-1264.

Stuster, J. (2001). *Albuquerque Police Department: Safe streets program.* (Report No. DOT HS 809 278).

Traffic Analysis Zone (TAZ) (n.d.). In *Census Glossary online.* Retrieved from https://www.census.gov/glossary/#term_TrafficAnalysisZoneTAZ

Transport Canada. (2011). Road safety in Canada. Retrieved from http://www.tc.gc.ca/eng/motorvehiclesafety/tp-tp15145-1201.htm

Transport Canada. (2013). Canadian Motor Vehicle Traffic Collision Statistics. Retrieved from https://www.tc.gc.ca/media/documents/roadsafety/cmvtcs2013_eng.pdf

Urban Planning Division, Community Services Department, City of Regina. (1992). Zoning Bylaw No. 9250.

Usman, T., Fu, L., & Miranda-Moreno, L. (2011). Accident Prediction Models for Winter Road Safety: Does Temporal Aggregation of Data Matter? *Transportation Research Record, Journal of the Transportation Research Board, 2011, 2237,* 144-151.

United States Department of Transportation, Federal Highway Administration. (2007). Defining Traffic Analysis Zones. Retrieved from https://www.fhwa.dot.gov/planning/tmip/publications/other_reports/technical_synthesis_report/page01.cfm

Van Schalwayk, I. (2000). Macro-level evaluation of road safety improvement interventions: An evaluation of the arrive alive I (1997/98) road safety campaign (Masters dissertation). University of Pretoria.

Van Schalwayk, I. (2008). The development of aggregated macro-level safety prediction models (Doctoral dissertation). Arizona State University.

Wallace, M., Turner, J., Matarazzo, A., & Babyak, C. (2009). Measuring crime in Canada: Introducing the crime severity index and improvements to uniform crime reporting survey. Ottawa, ON: Statistics Canada.

Washington, S., Persuad, B., Lyon, C., & Oh, J. (2005). Validation of accident models for intersections. *Federal Highway Administration, U.S Department of Transportation.*

Weisburd, D. (2008). Ideas in American Policing: Place-based policing. *Police Foundation, 2008,* 9.

Weisburd, D, L., Groff, E., & Morris, N. (2011). Hotspots of Juvenile Crime: findings from Seattle. *Juvenile Justice Bulletin, Office of Juvenile Justice and delinquency Prevention.* Retrieved from https://www.ncjrs.gov/pdffiles1/ojjdp/231575.pdf

Wheeler, A. P. (2014). Visualization techniques for journey to crime flow data. *Cartography and Geographic Information Science, 2015, 42*(2), 149-161.

Wheeler, A. P. (2015). What we can learn from small units of analysis (Doctoral dissertation). University at Albany, State University of New York.

World Health Organization. (2006). Road traffic injury prevention: Training manual. Dehli, India: Indian Institute of Technology.

Yannis, G., Papadimitriou, E., & Antoniou, C. (2007). Multilevel modeling for the regional effect of enforcement on road accidents. *Accident Analysis & Prevention, 39,* 818-825.

Yu, Liguo. (2012). Using negative binomial regression analysis to predict software faults: A study of Apache Ant. *International Journal of Information Technology and Computer Science, 8*, 63-70.

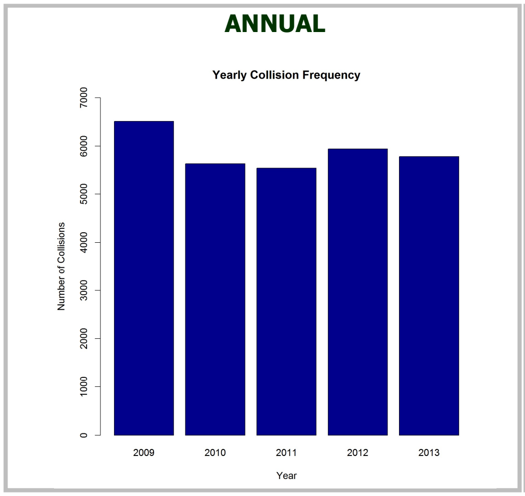
Ziad, A., Sawalha. (2002). Traffic accident modeling: Statistical issues and safety application (Doctoral dissertation). University of British Columbia, Vancouver.

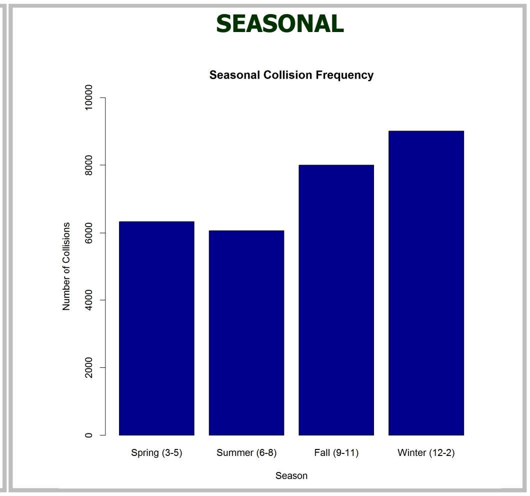
Zegeer, C. V. (1982) *Highway collision analysis systems*, NCHRP, 91, TRB, National Cooperative Highway Research Program, Washington D.C.

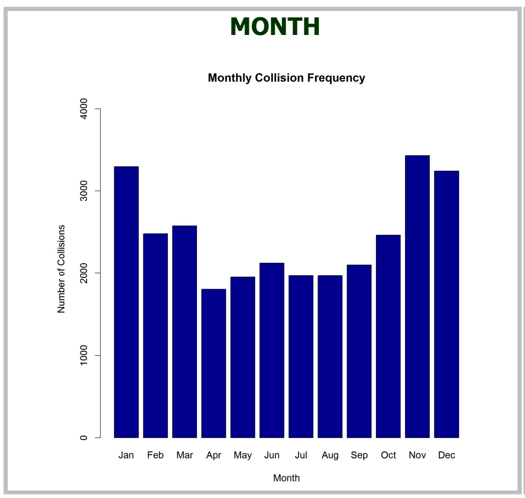
APPENDICES

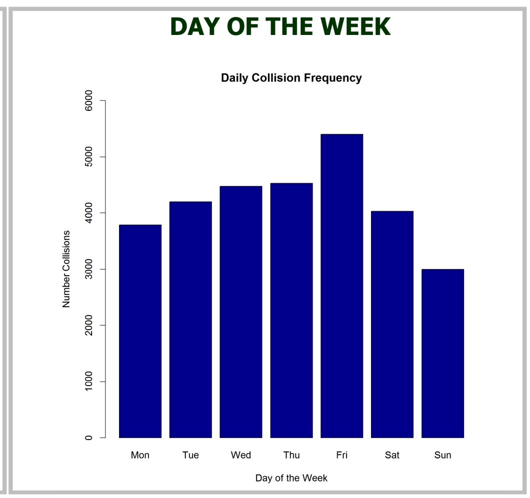
APPENDIX A: Temporal Descriptive Statistics of Data

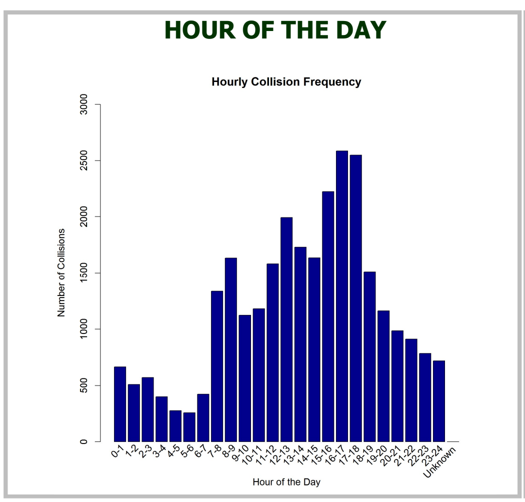
A1. Total Collisions Bar Charts: Collison Frequency by Year, Season, Month, Day of the Week, Hour of the Day, and Per Capita.

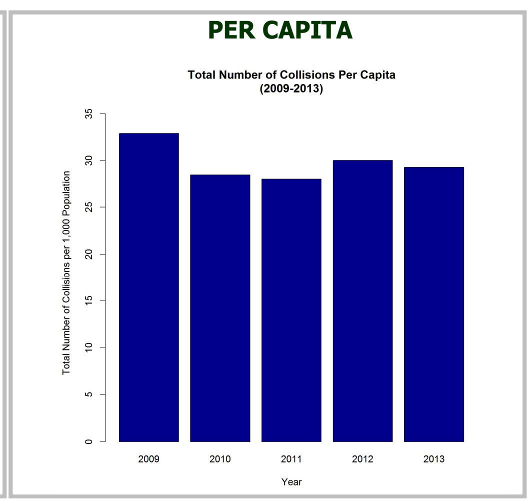




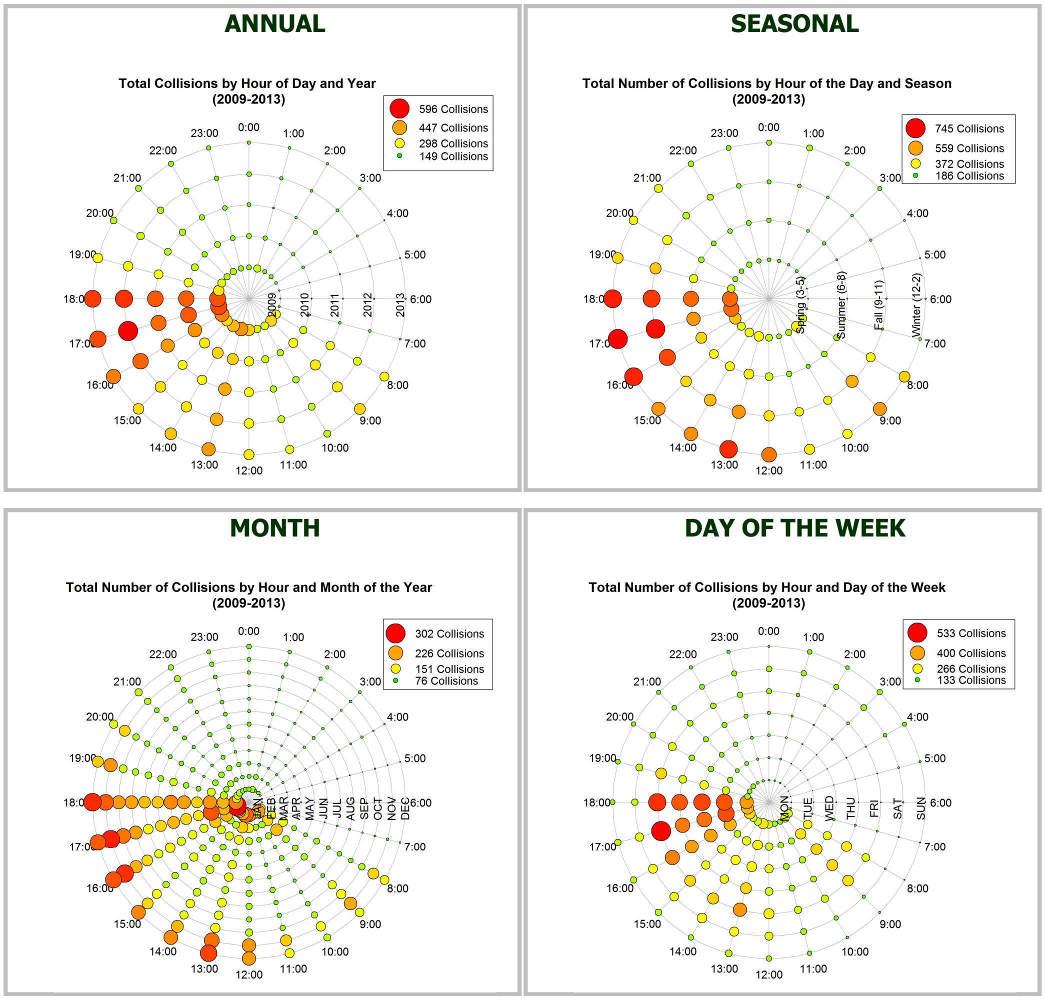




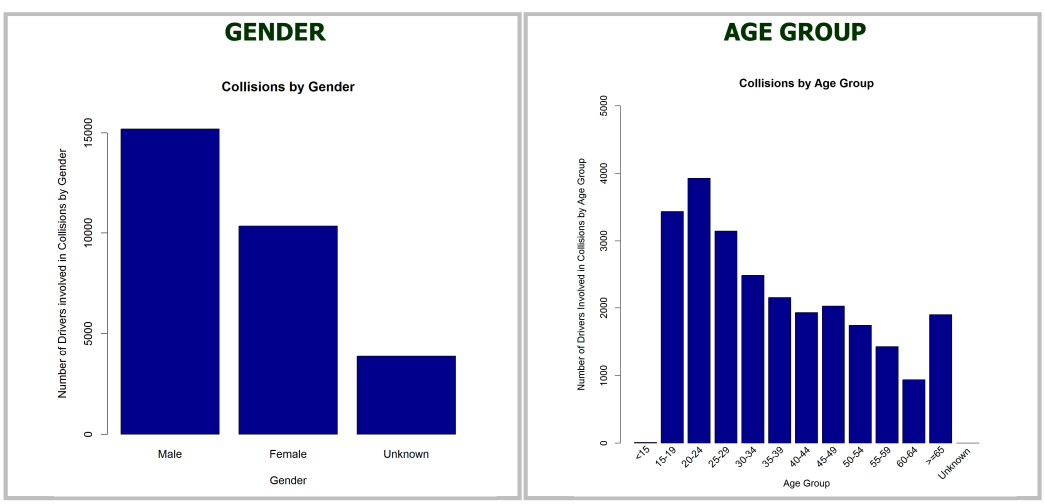




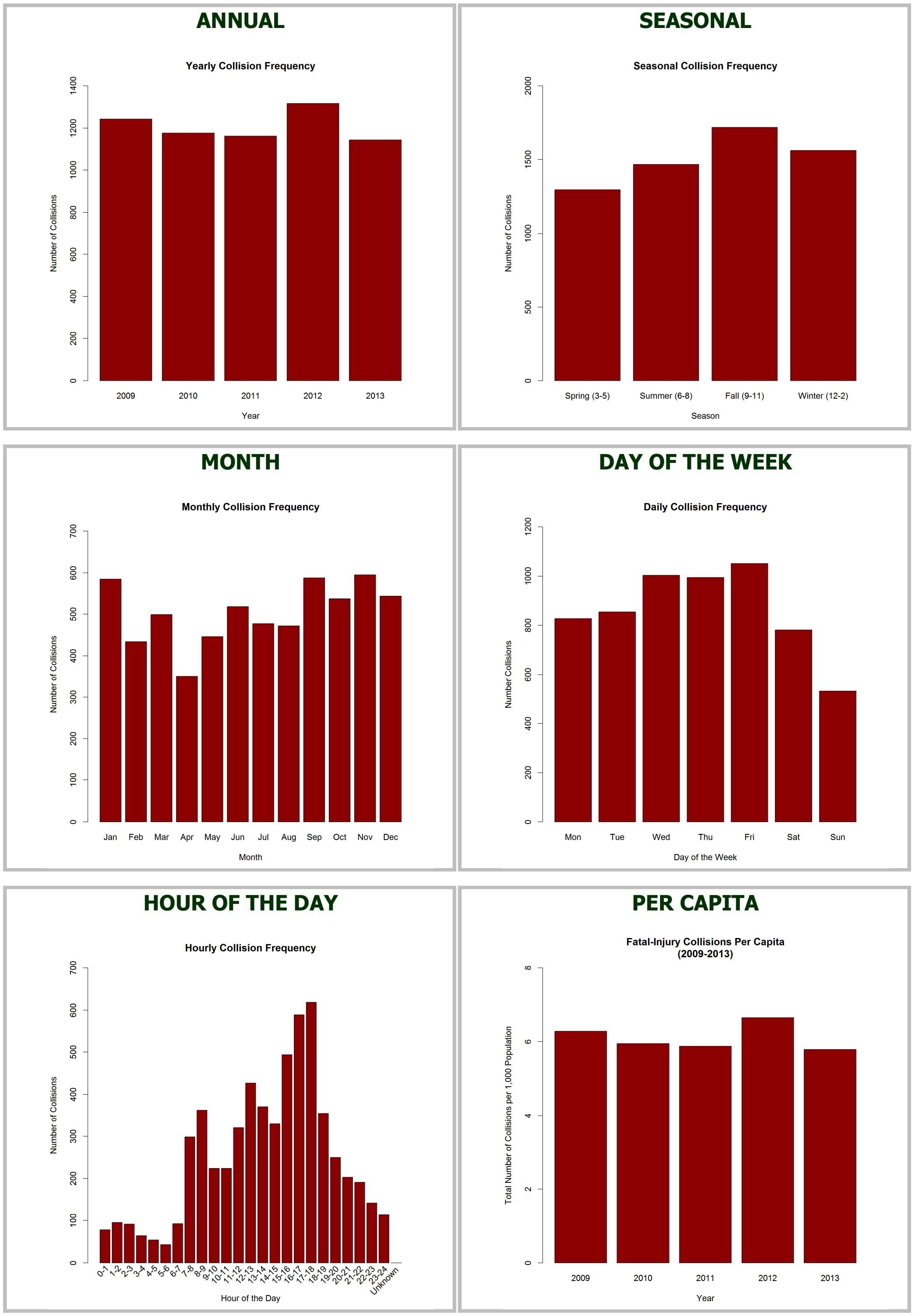
A2. Total Collisions Clockplots: Collison Frequency by Year, Season, Month, and Day of the Week.



A3. Total Collisions Bar Charts: Collison Frequency by Gender and Age Group

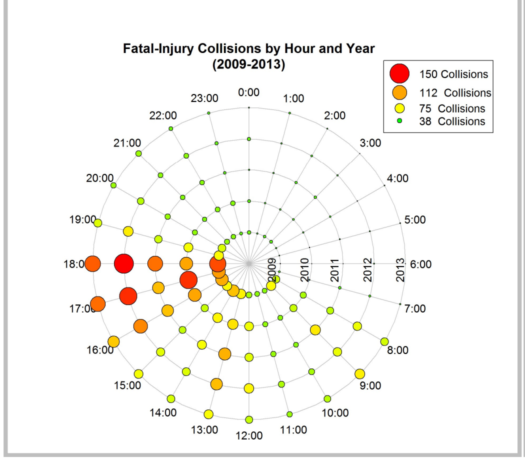


A4. Fatal-Injury Collisions Bar Charts: Collison Frequency by Year, Season, Month, Day of the Week, Hour of the Day, and Per Capita.

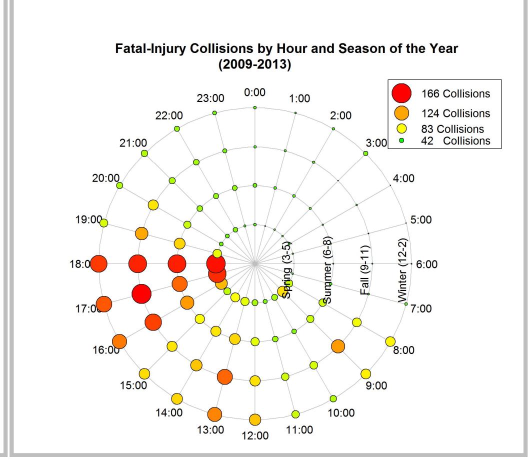


A5. Fatal-Injury Collisions Clockplots: Collison Frequency by Year, Season, Month, and Day of the Week.

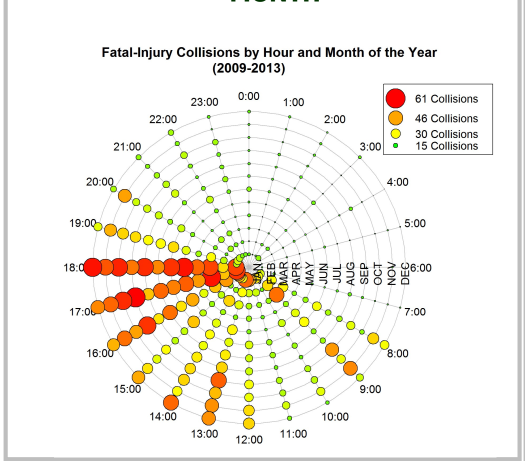
ANNUAL



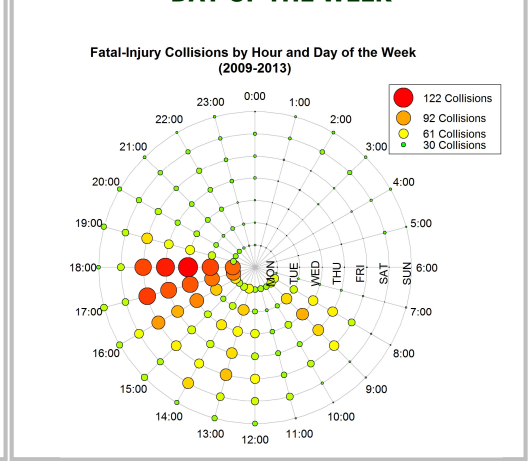
SEASONAL



MONTH

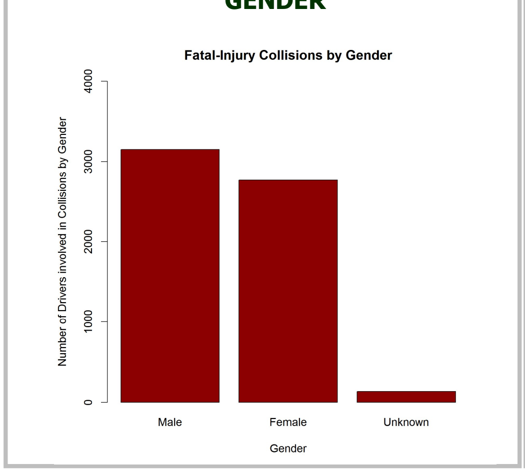


DAY OF THE WEEK

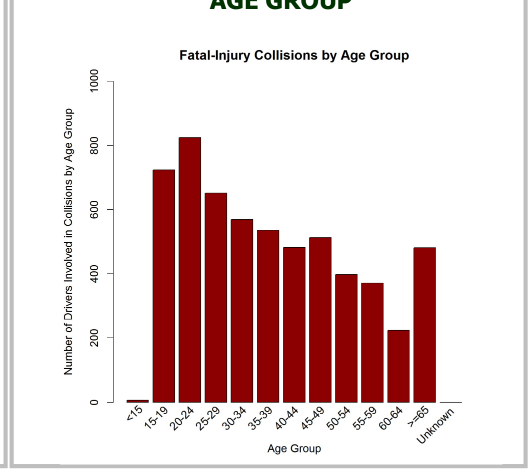


A6. Fatal-Injury Collisions Bar Charts: Collison Frequency by Gender and Age Group

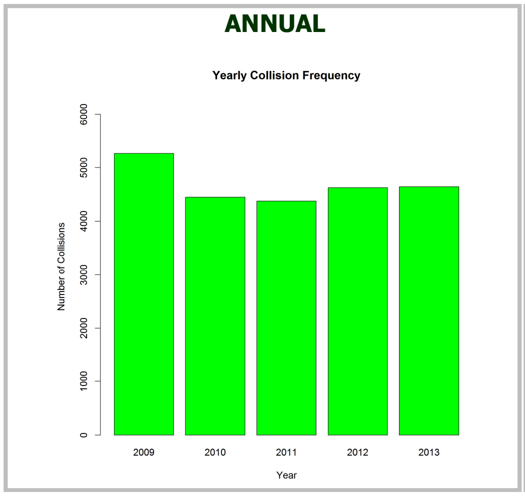
GENDER

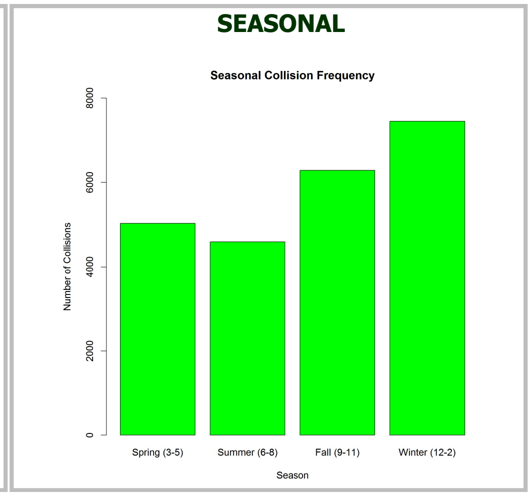


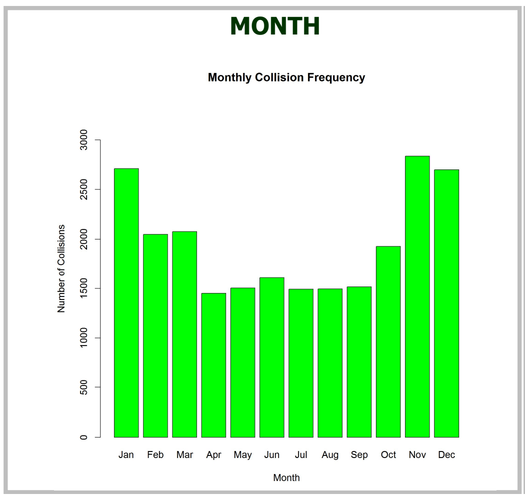
AGE GROUP

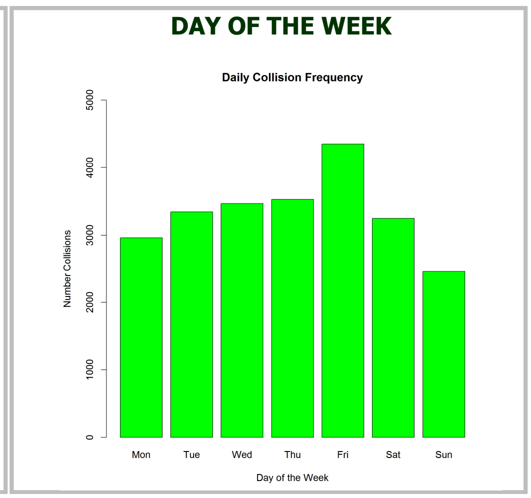


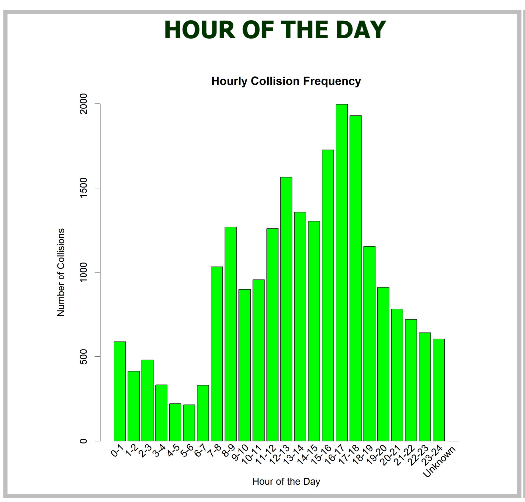
A7. Property Damage Only Collisions Bar Charts: Collisions Bar Charts: Collison Frequency by Year, Season, Month, Day of the Week, Hour of the Day, and Per Capita.

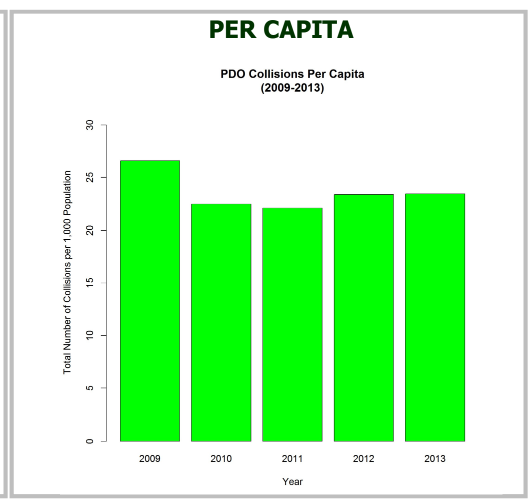




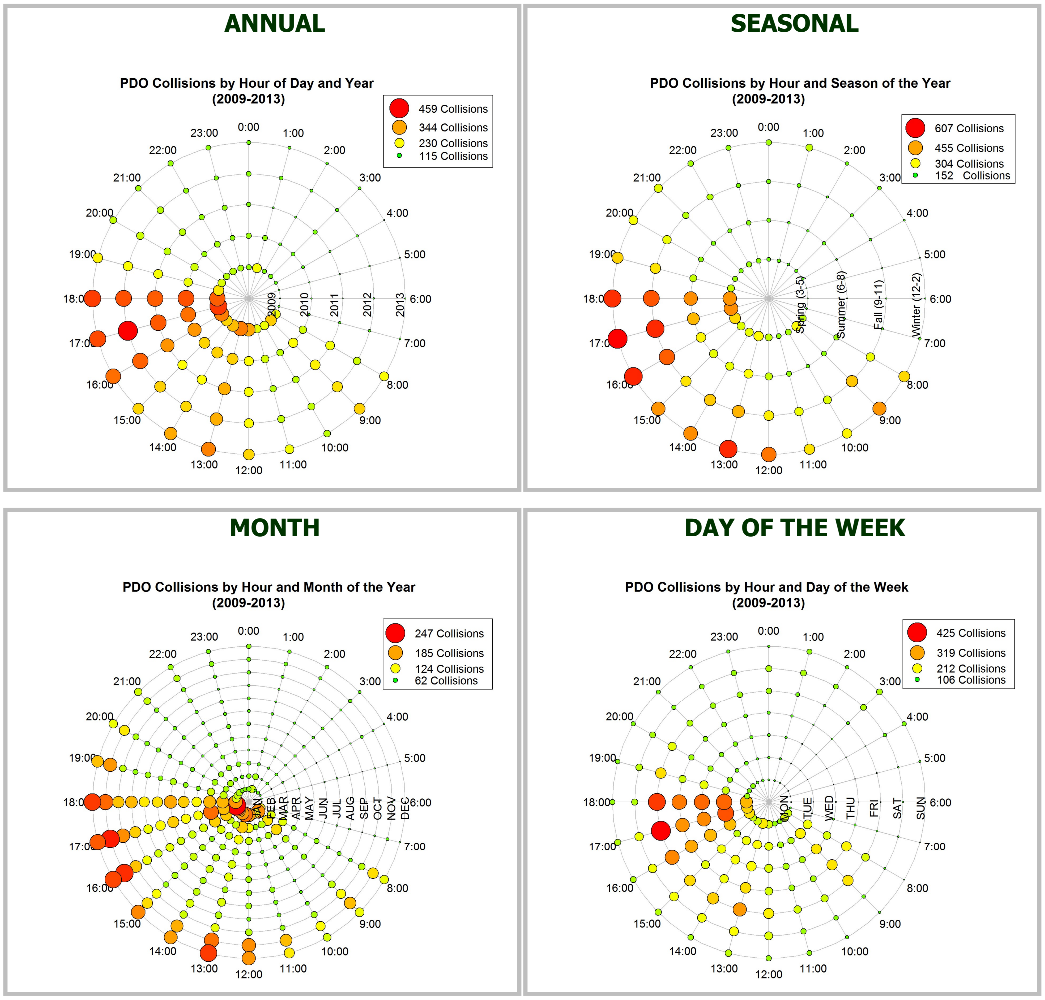




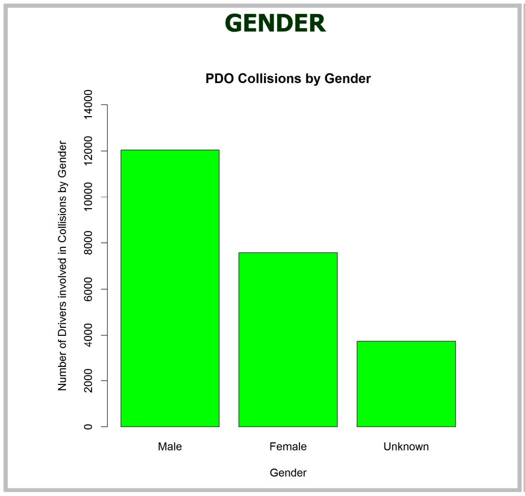


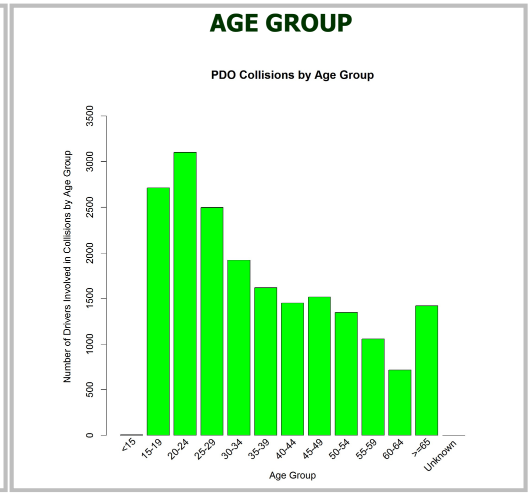


A8. Property Damage Only Collisions Clockplots: Collison Frequency by Year, Season, Month, and Day of the Week.

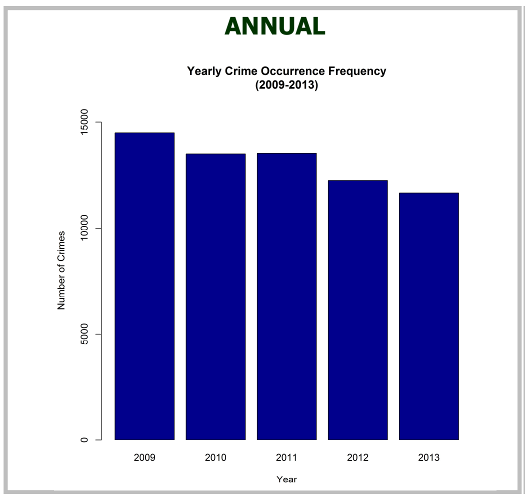


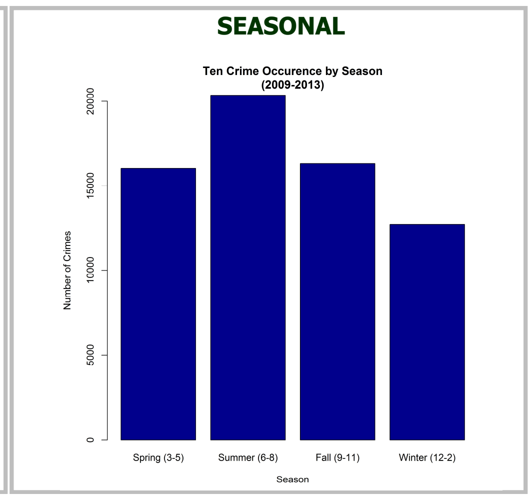
A9. Property Damage Only Collisions Bar Charts: Collison Frequency by Gender and Age Group

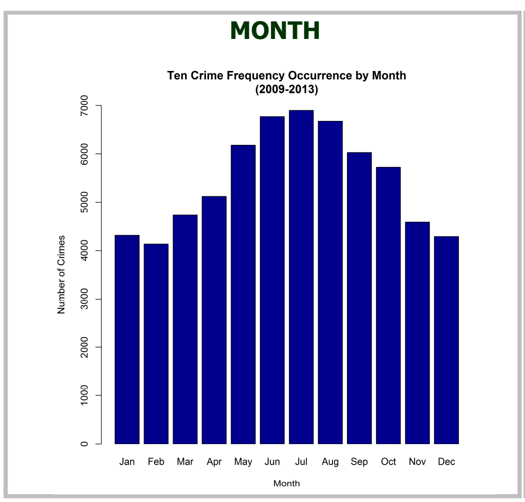


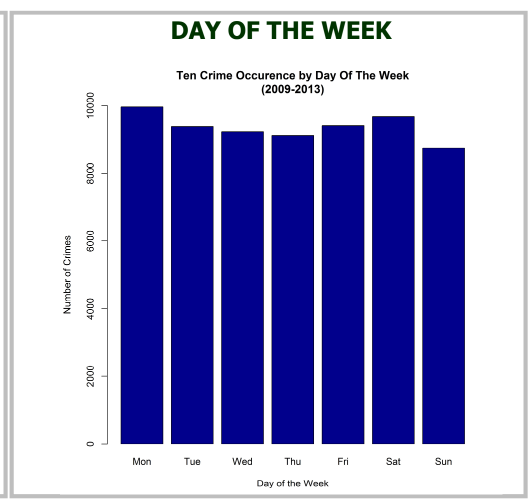


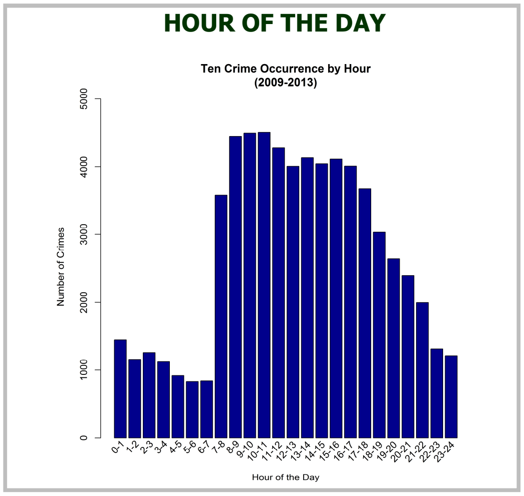
A10. Total Crimes Bar Charts: Collisions Bar Charts: Collison Frequency by Year, Season, Month, Day of the Week, and Hour of the Day.



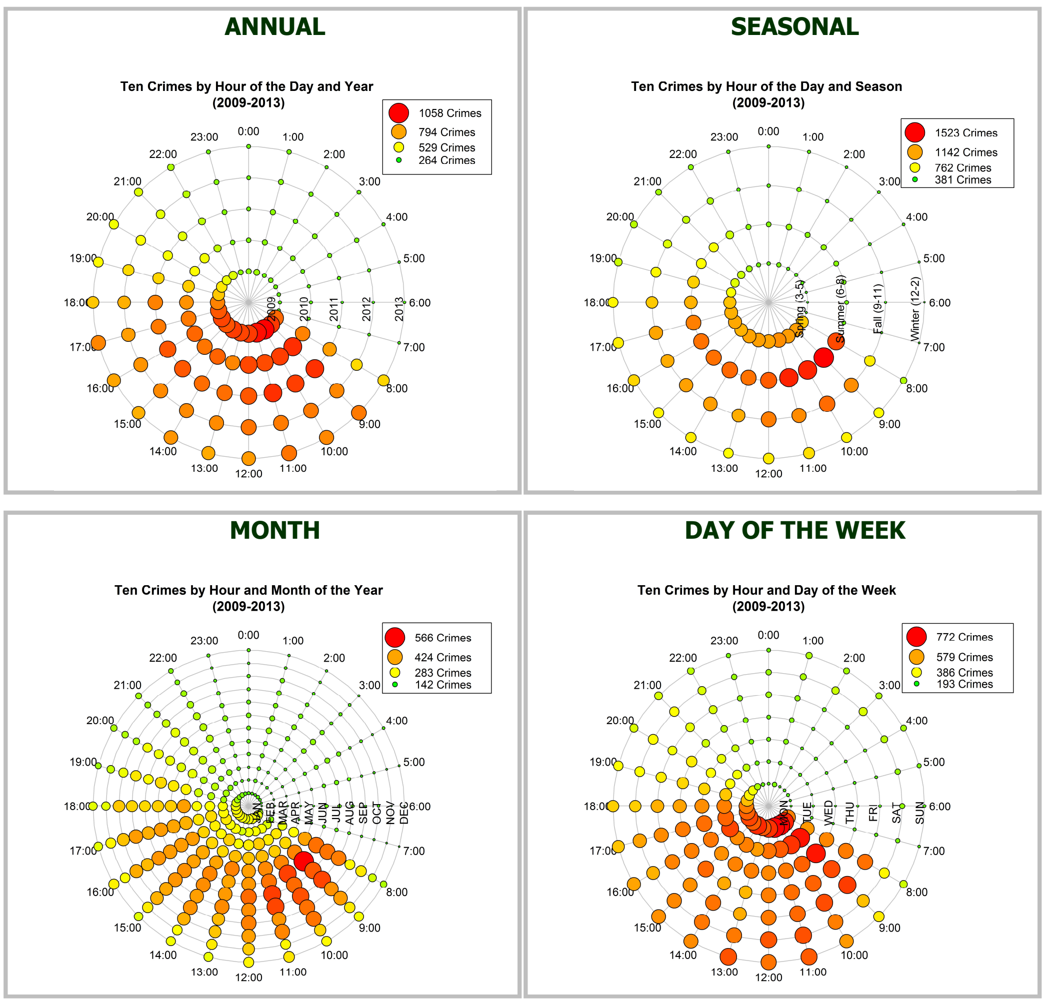




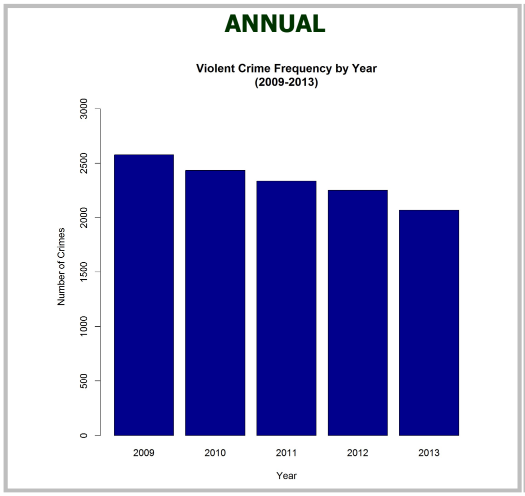


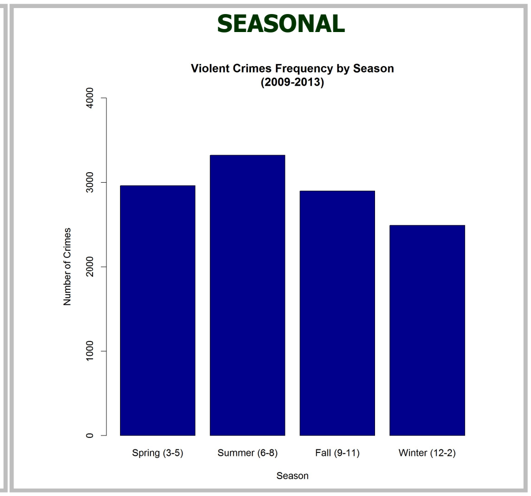


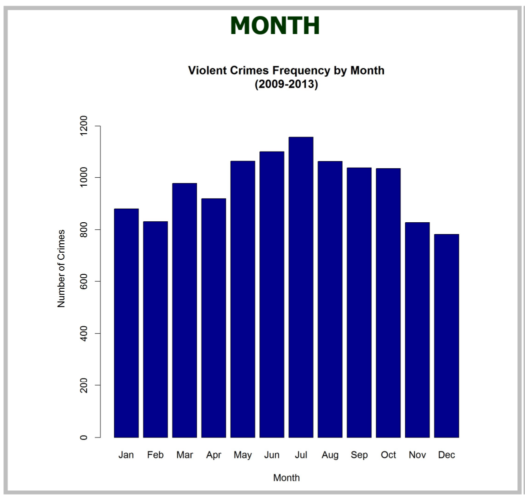
A11. Total Crimes Clockplots: Collison Frequency by Year, Season, Month, and Day of the Week.

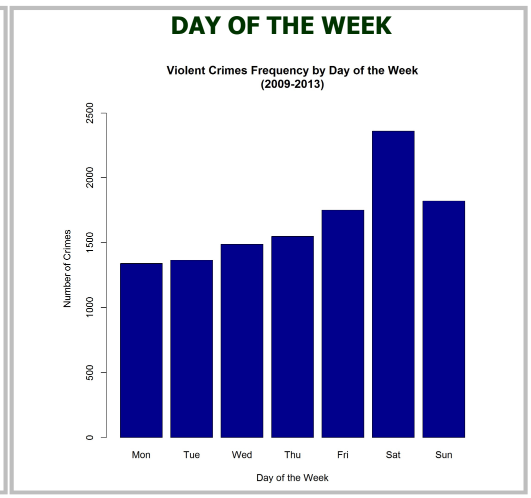


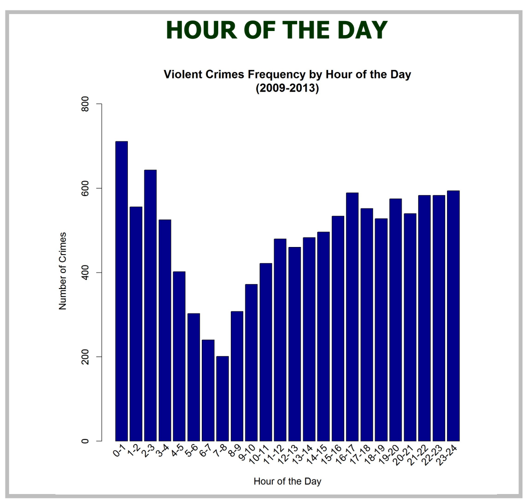
A12. Violent Crimes Bar Charts: Collisions Bar Charts: Collison Frequency by Year, Season, Month, Day of the Week, and Hour of the Day.



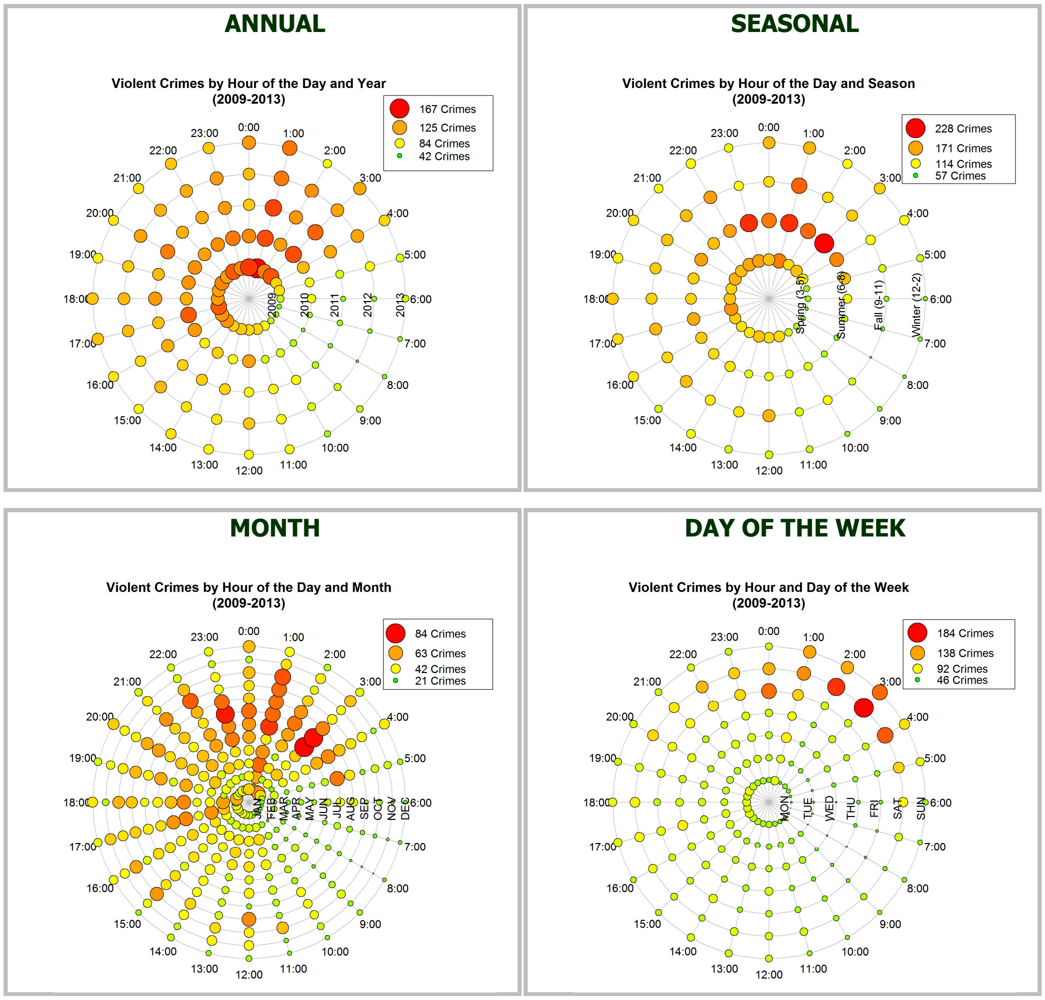




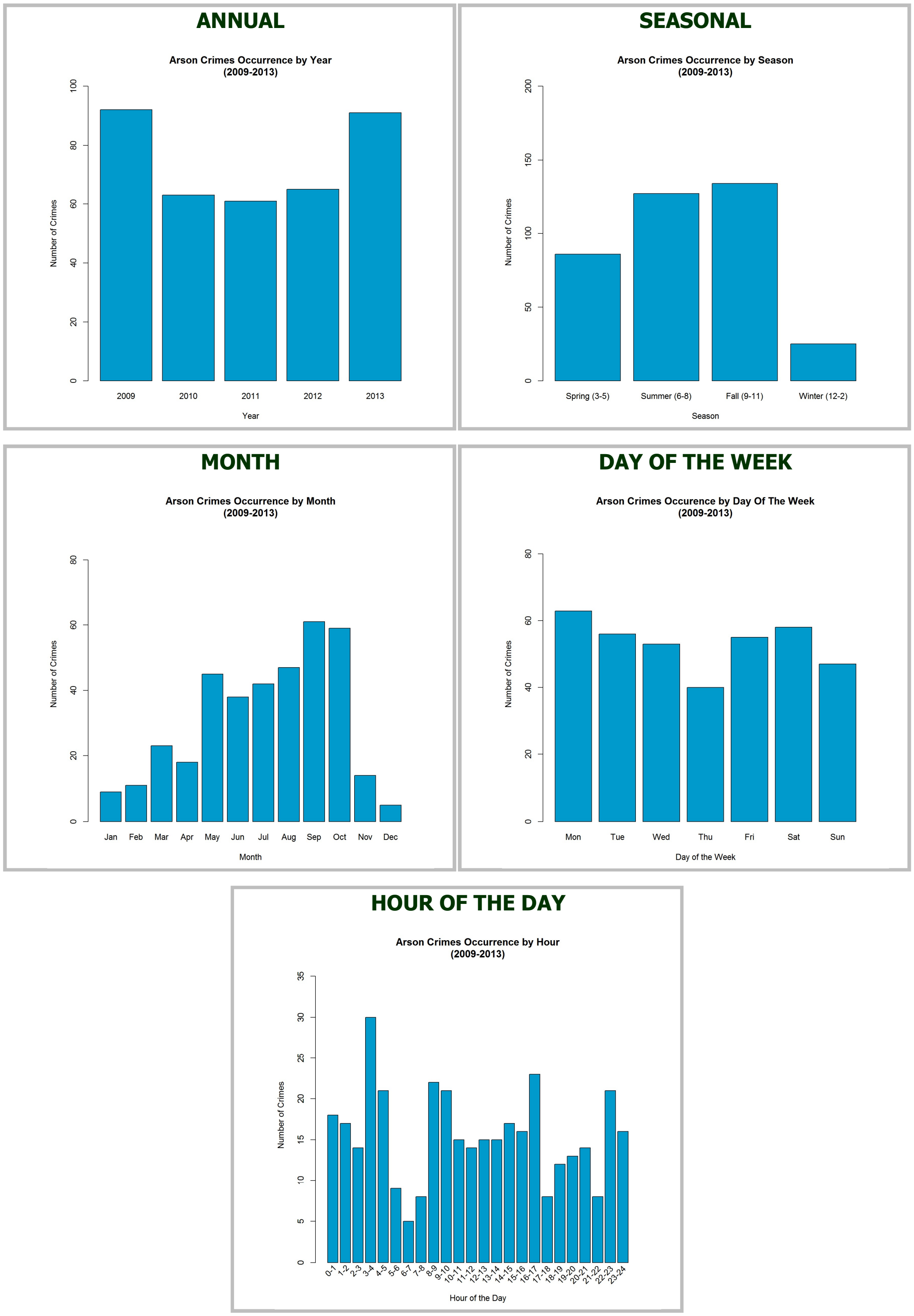




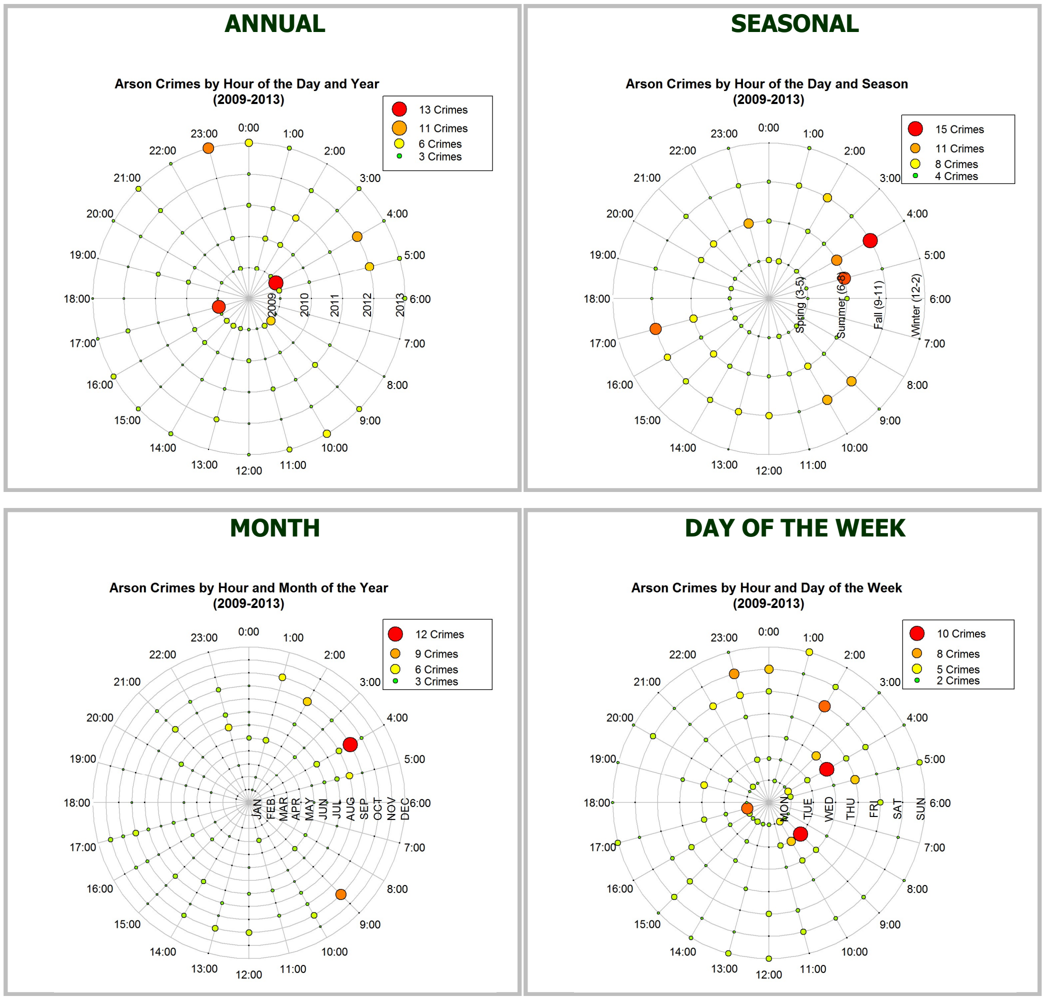
A13. Violent Crimes Clockplots: Collison Frequency by Year, Season, Month, and Day of the Week.



A14. Arson Crimes Bar Charts: Collisions Bar Charts: Collison Frequency by Year, Season, Month, Day of the Week, and Hour of the Day.

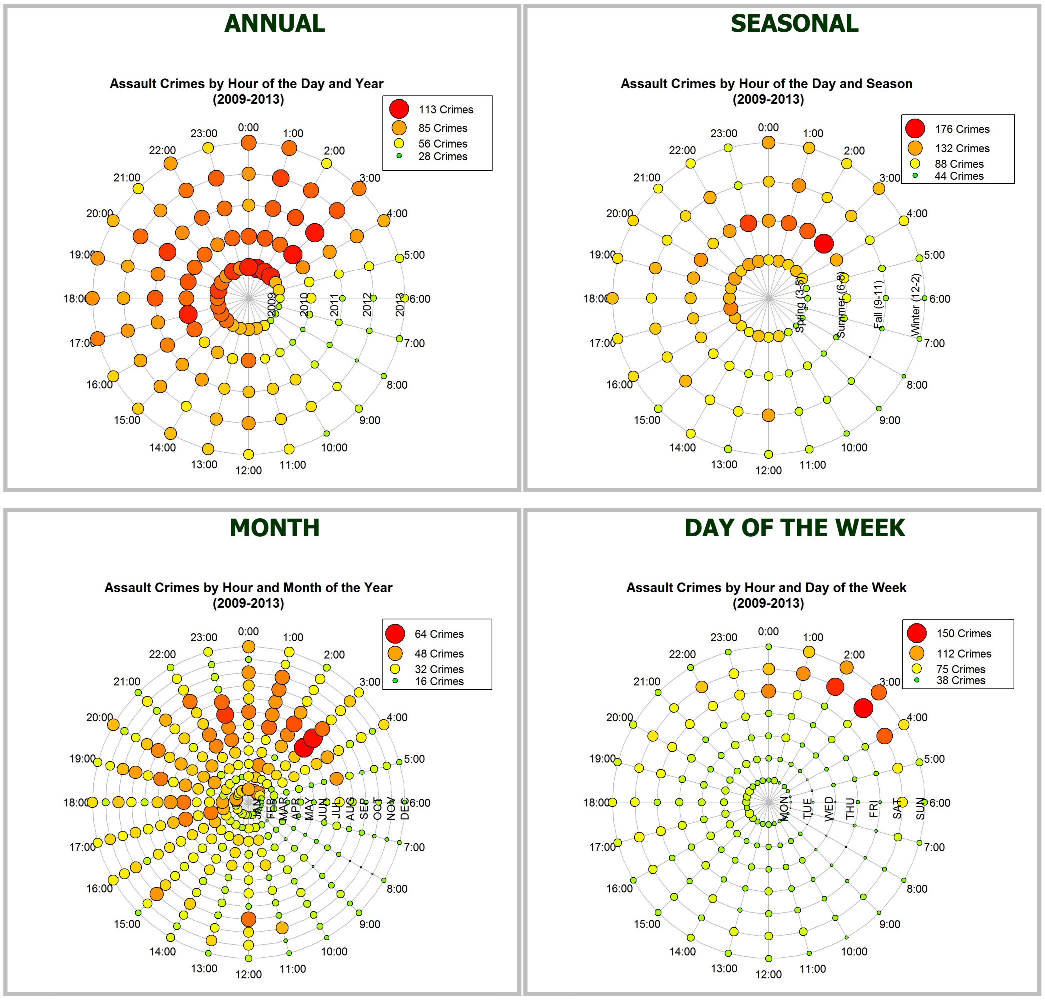


A15. Arson Crimes Clockplots: Collison Frequency by Year, Season, Month, and Day of the Week.

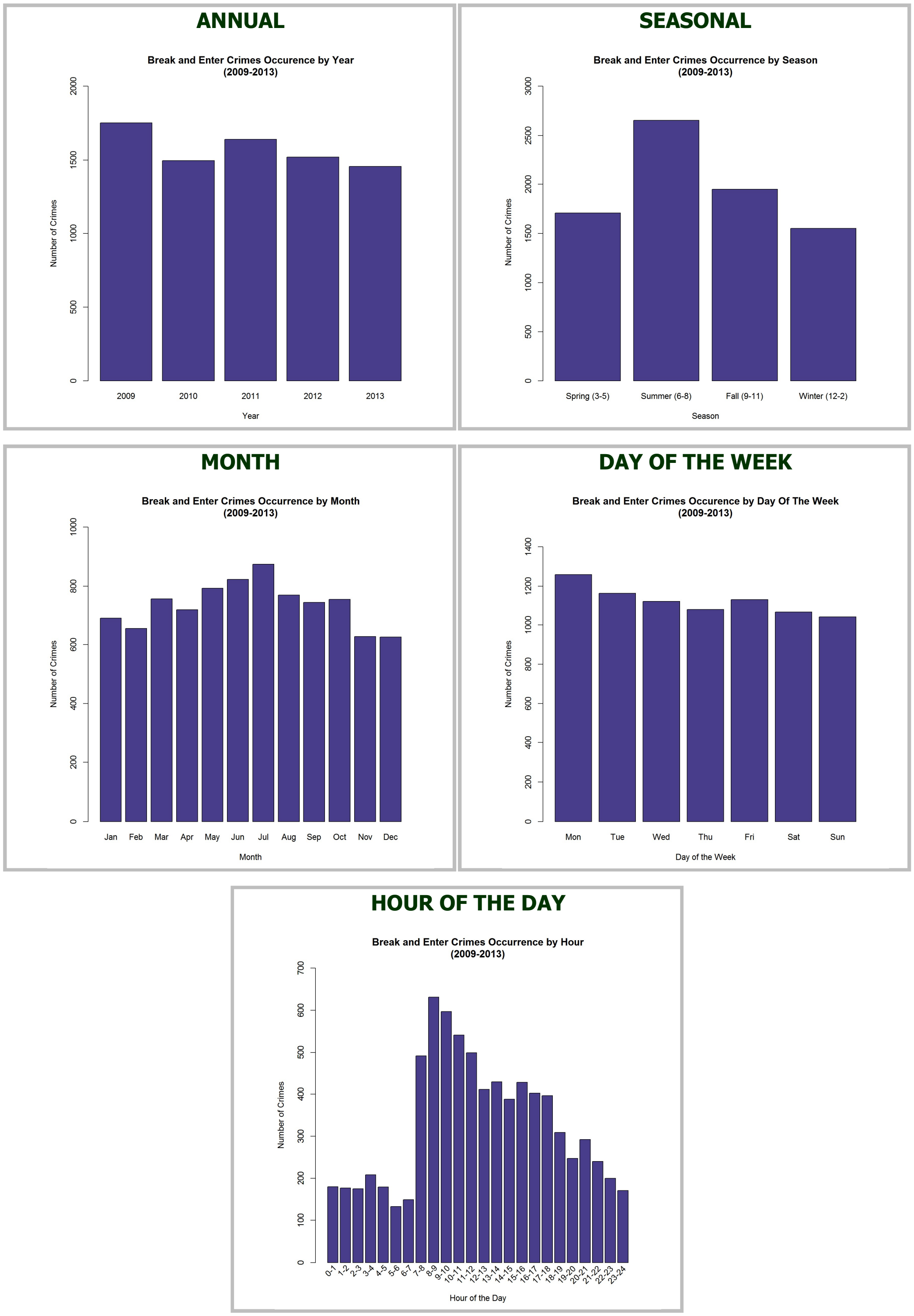


A16. Assault Crimes Bar Charts: Collisions Bar Charts: Collison Frequency by Year, Season, Month, Day of the Week, and Hour of the Day.

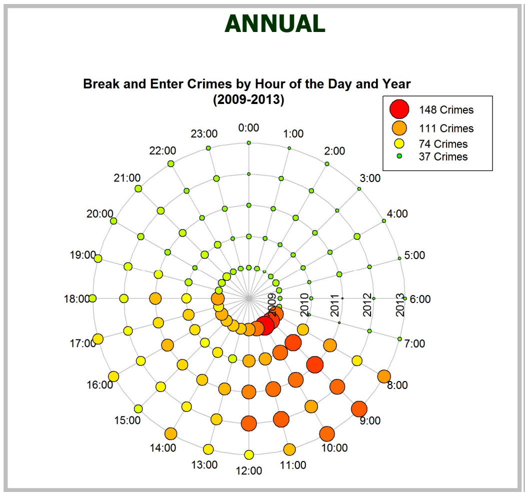
A17. Assault Crimes Clockplots: Collison Frequency by Year, Season, Month, and Day of the Week.

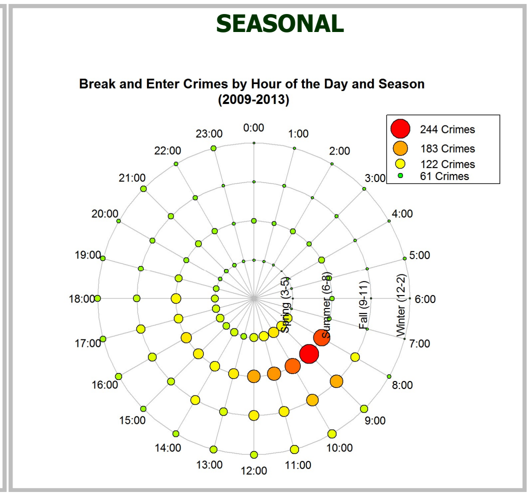


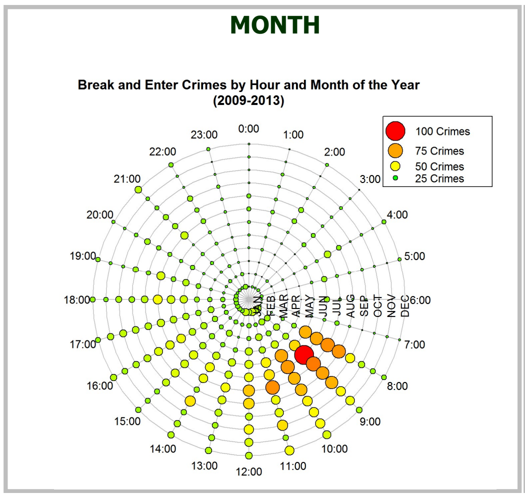
A18. Break and Enter Crimes Bar Charts: Collisions Bar Charts: Collison Frequency by Year, Season, Month, Day of the Week, and Hour of the Day.

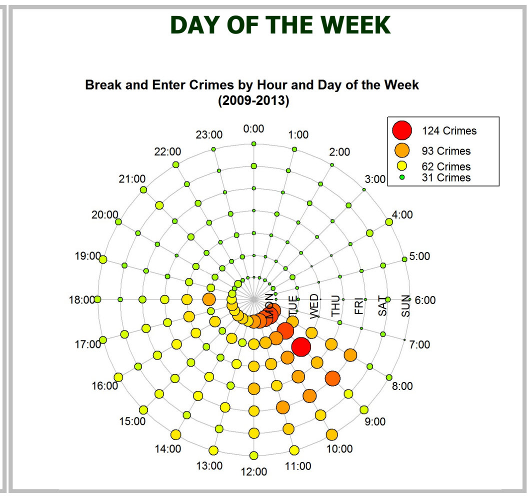


A19. Break and Enter Crimes Clockplots: Collison Frequency by Year, Season, Month, and Day of the Week.

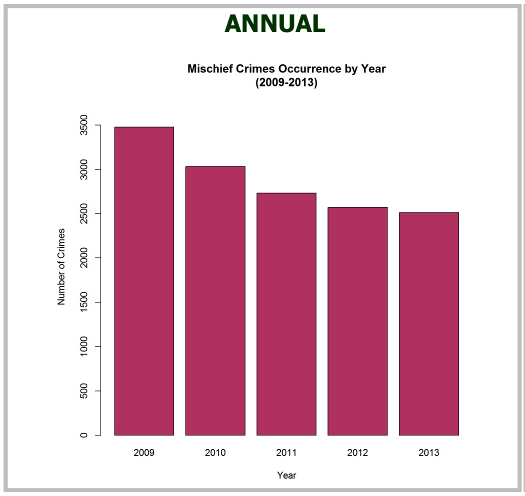


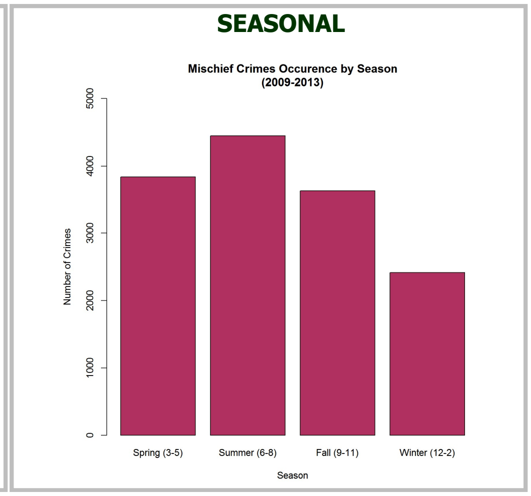


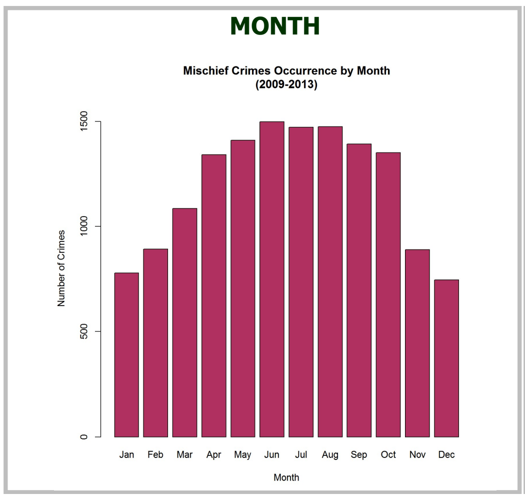


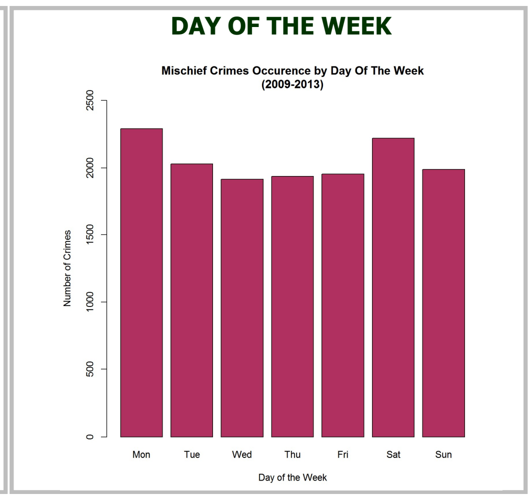


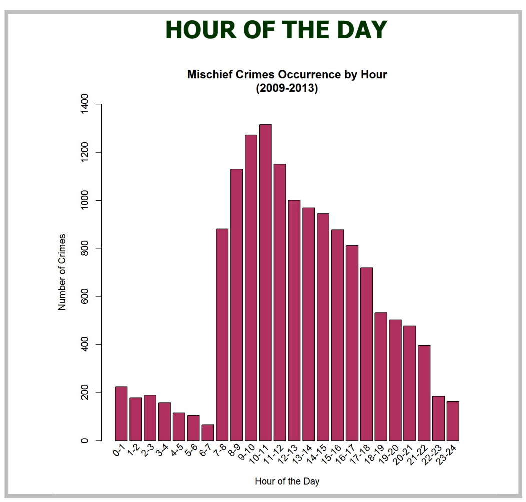
A20. Mischief Crimes Bar Charts: Collisions Bar Charts: Collison Frequency by Year, Season, Month, Day of the Week, and Hour of the Day.



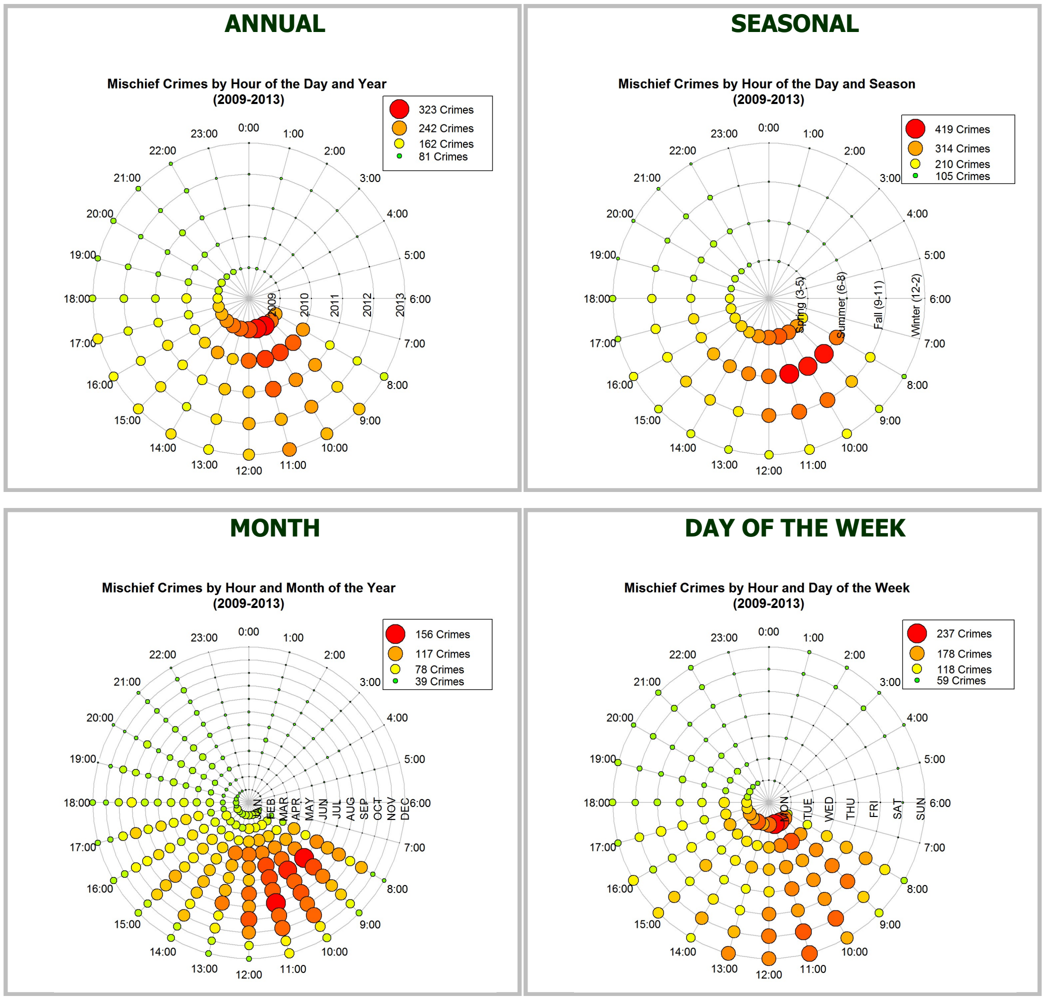




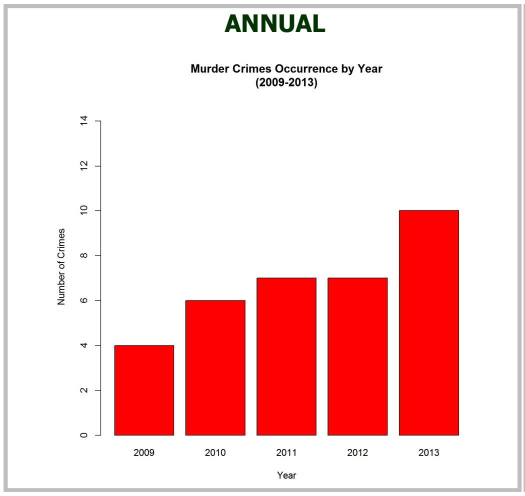


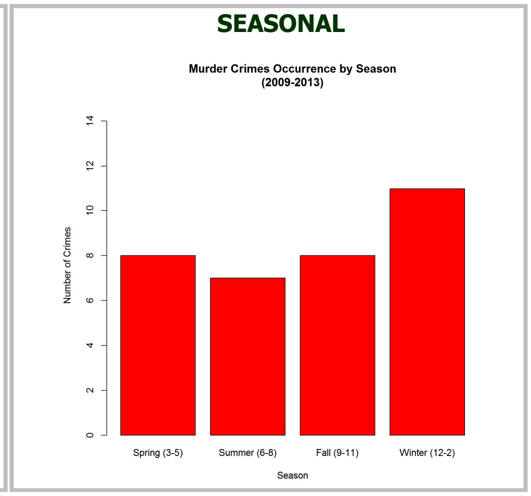


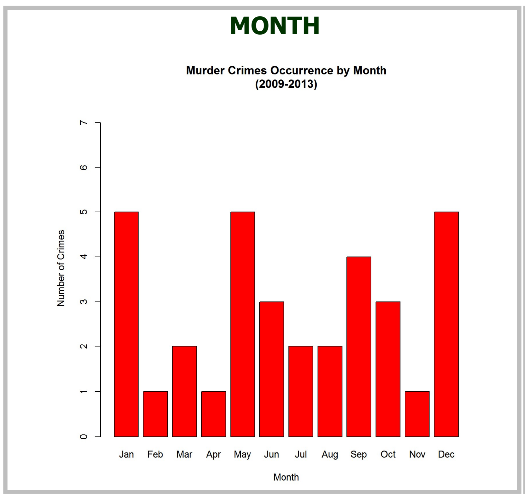
A21. Mischief Crimes Clockplots: Collison Frequency by Year, Season, Month, and Day of the Week.

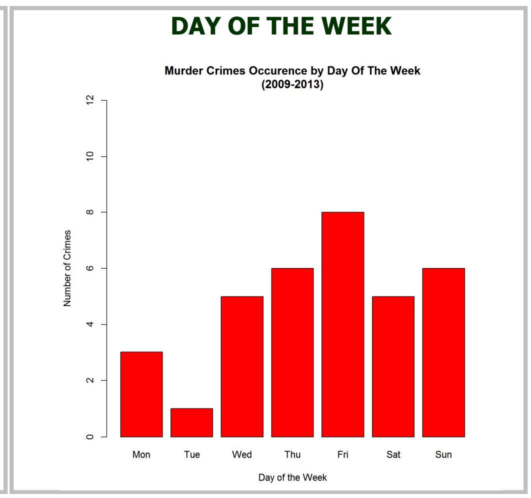


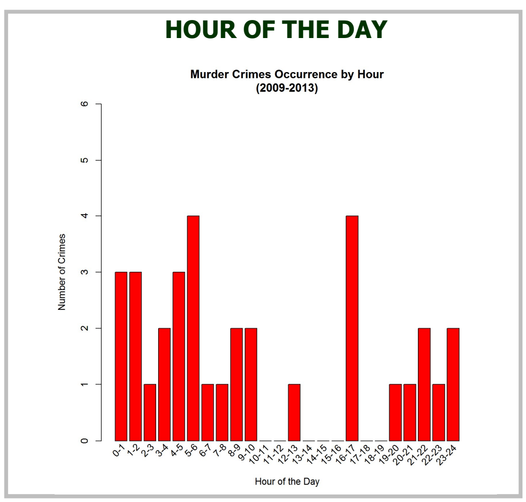
A22. Murder Crimes Bar Charts: Collisions Bar Charts: Collison Frequency by Year, Season, Month, Day of the Week, and Hour of the Day.



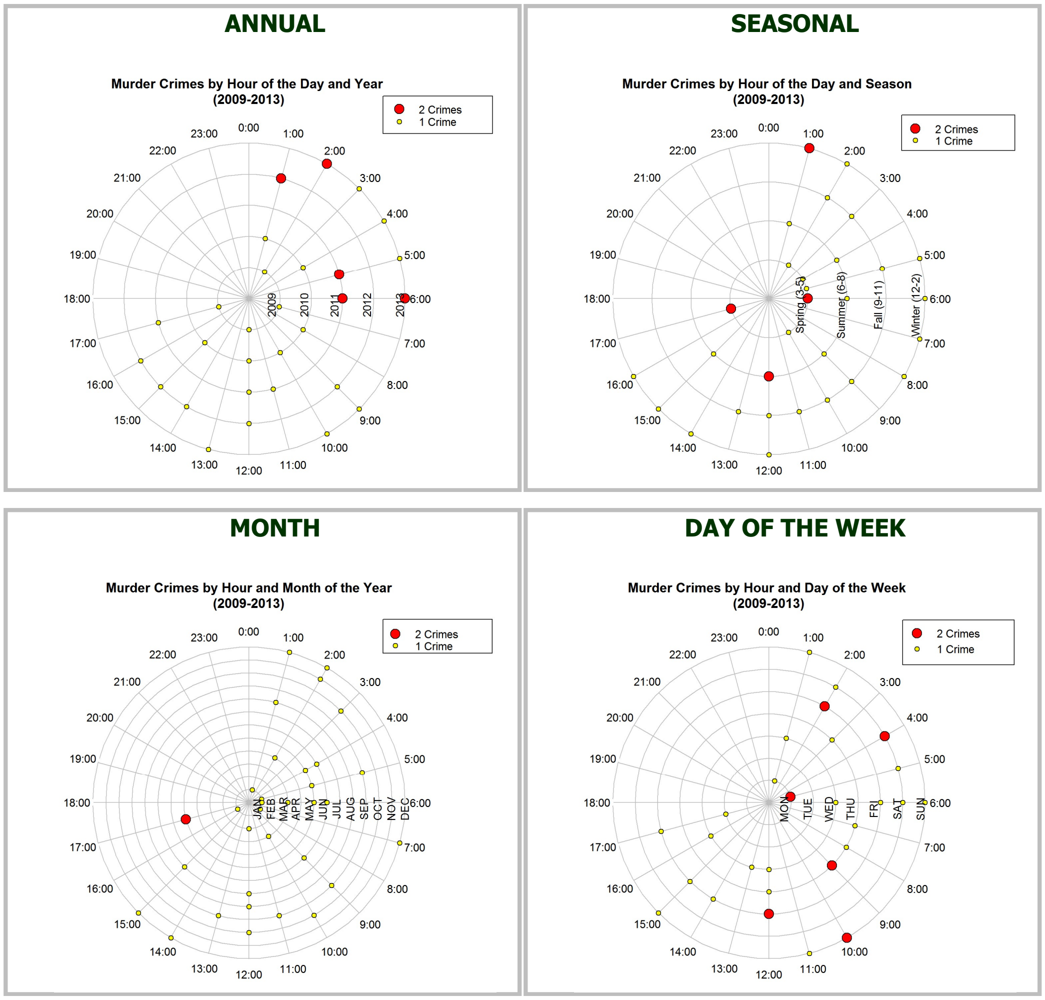




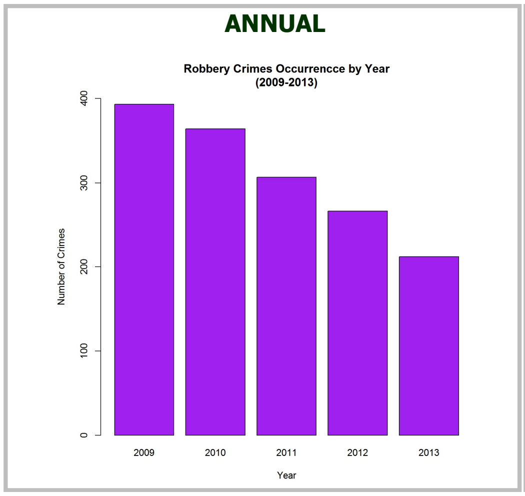


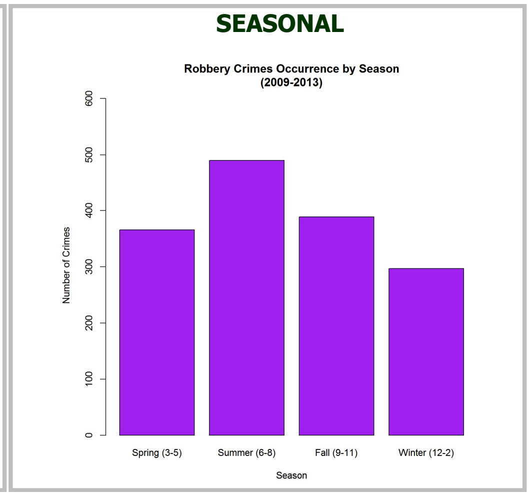


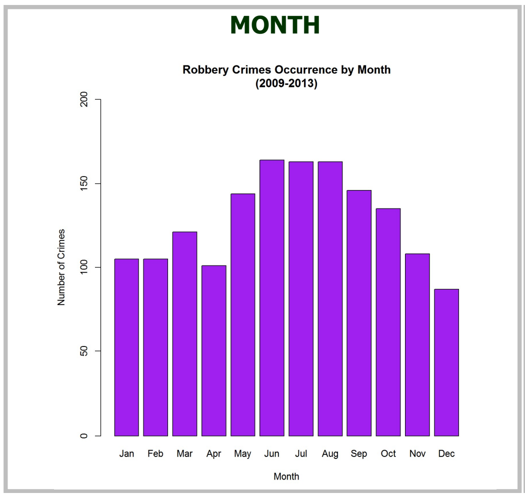
A23. Murder Crimes Clockplots: Collison Frequency by Year, Season, Month, and Day of the Week.

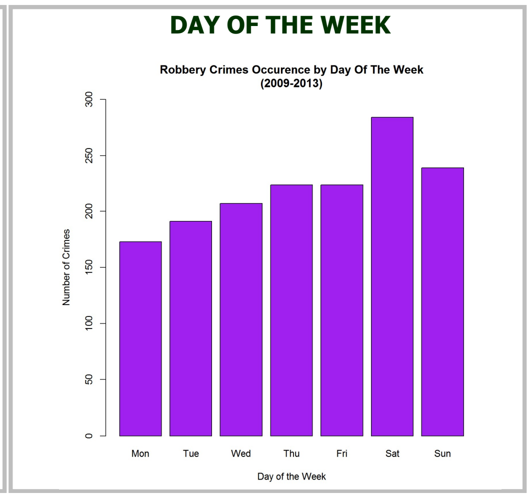


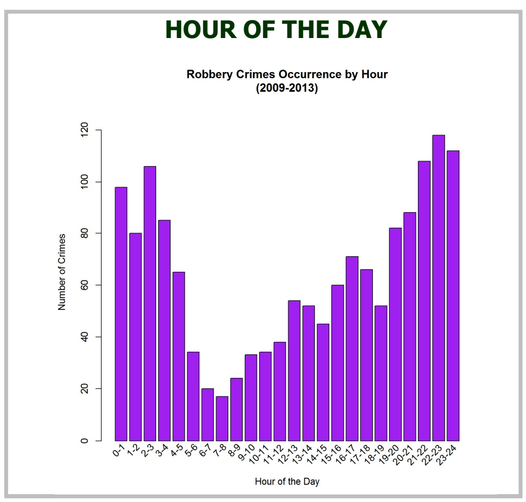
A24. Robbery Crimes Bar Charts: Collisions Bar Charts: Collison Frequency by Year, Season, Month, Day of the Week, and Hour of the Day.



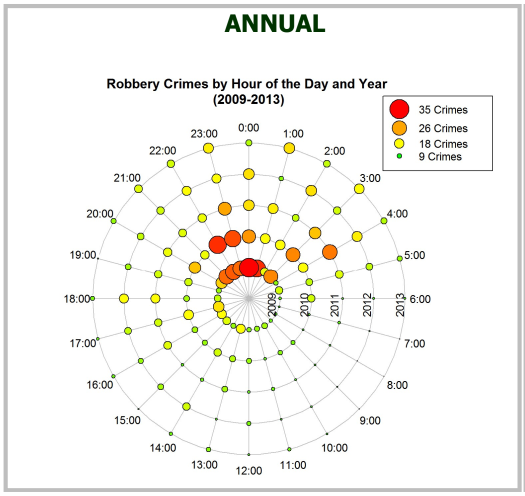


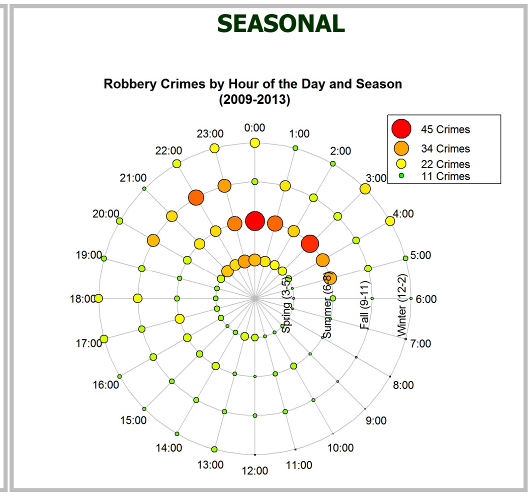


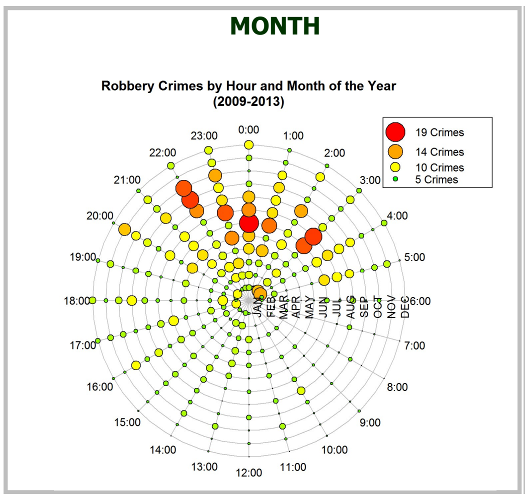


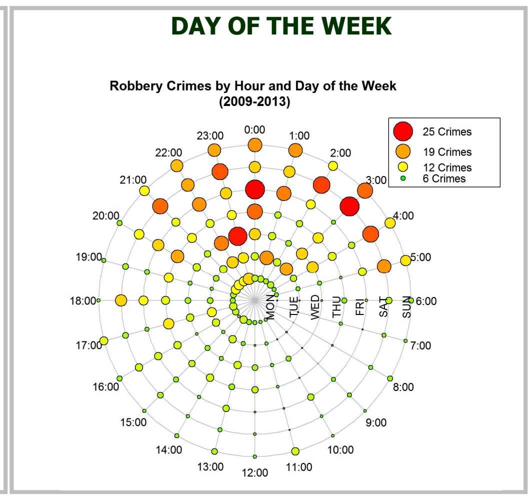


A25. Robbery Crimes Clockplots: Collison Frequency by Year, Season, Month, and Day of the Week.

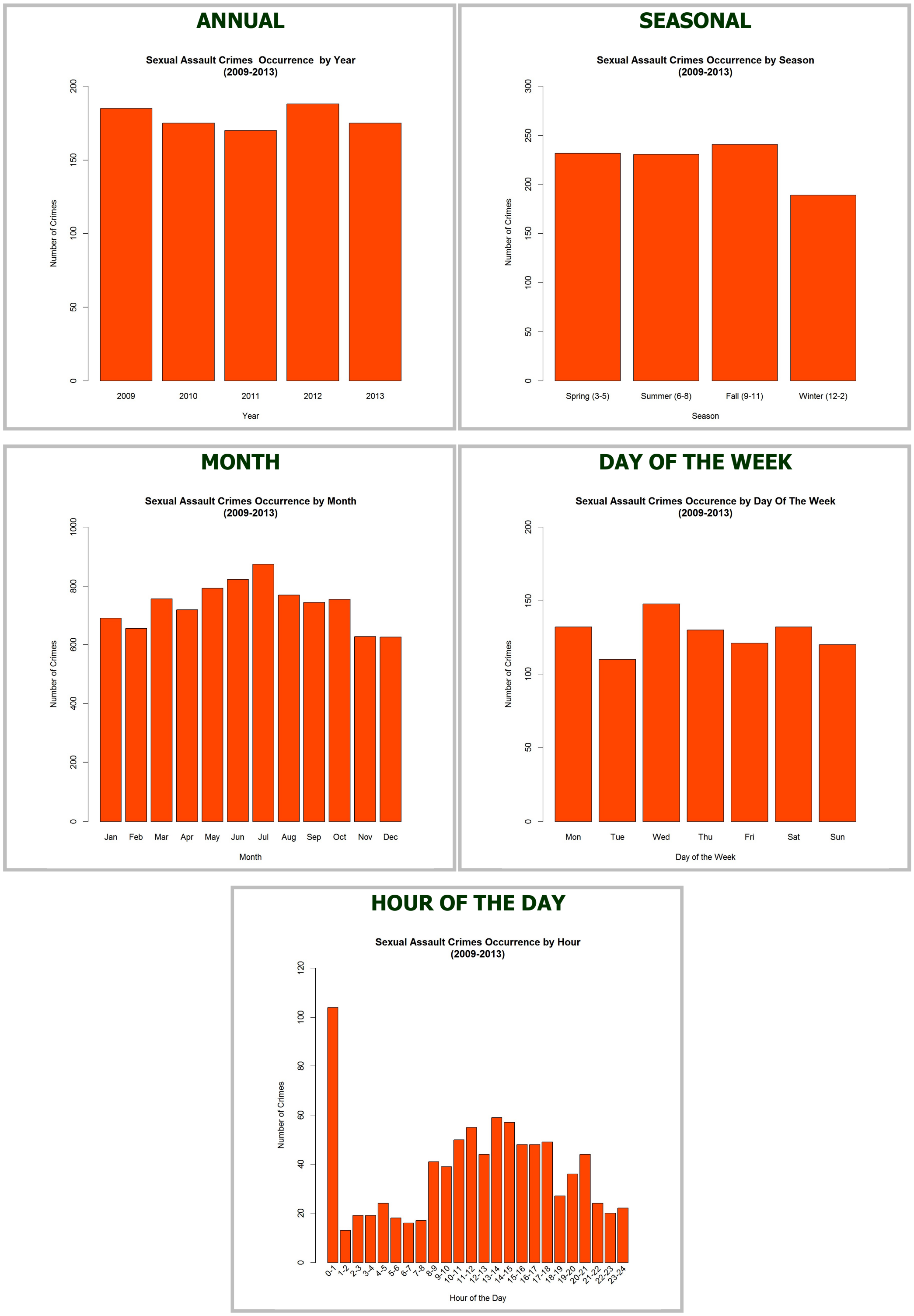




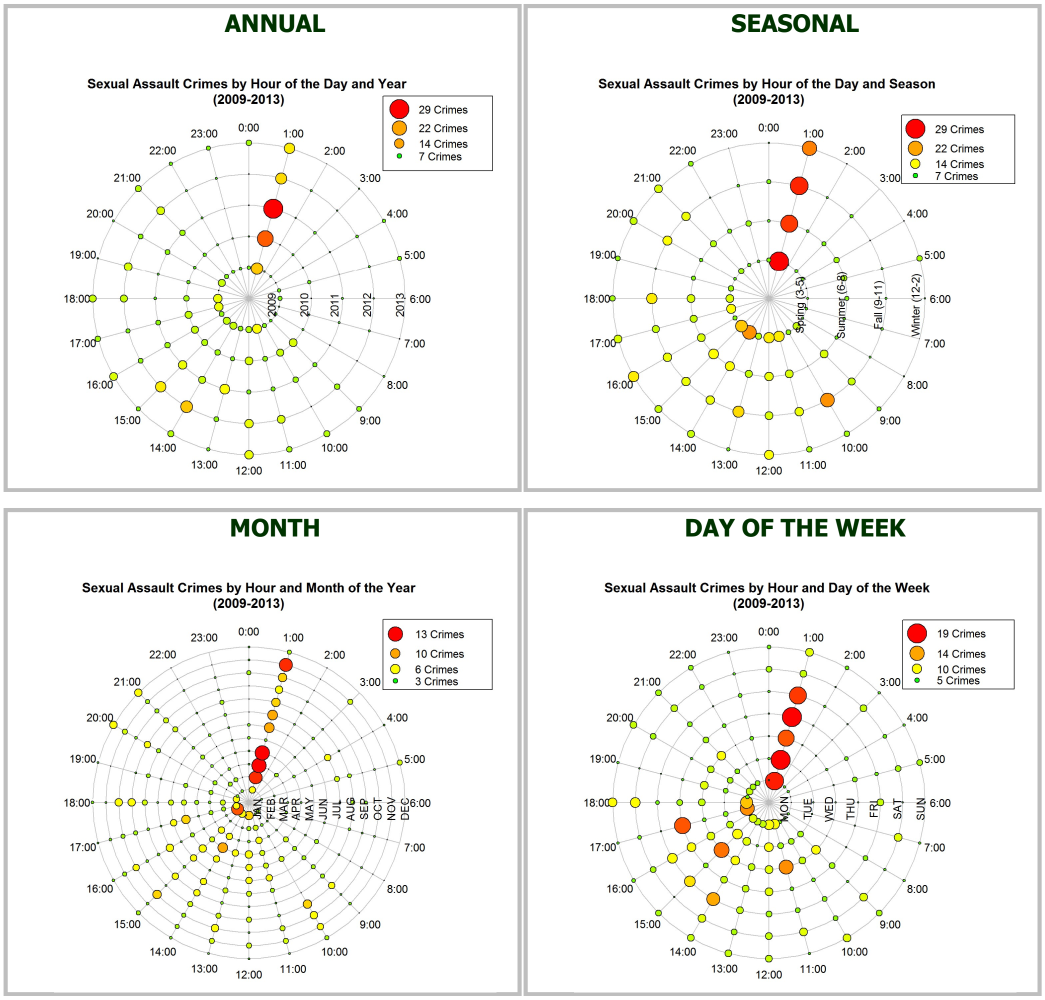




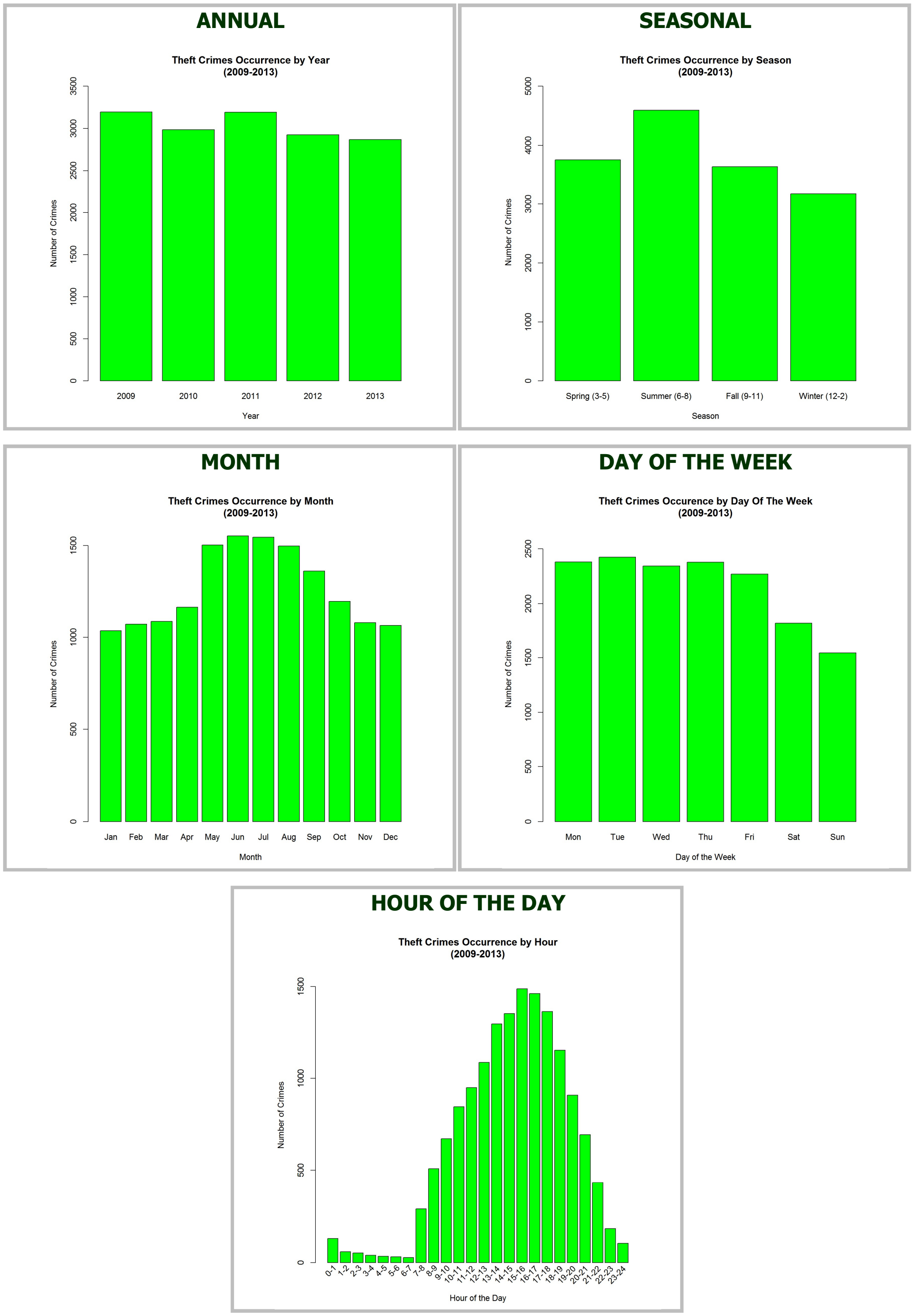
A26. Sexual Assault Crimes Bar Charts: Collisions Bar Charts: Collison Frequency by Year, Season, Month, Day of the Week, and Hour of the Day.



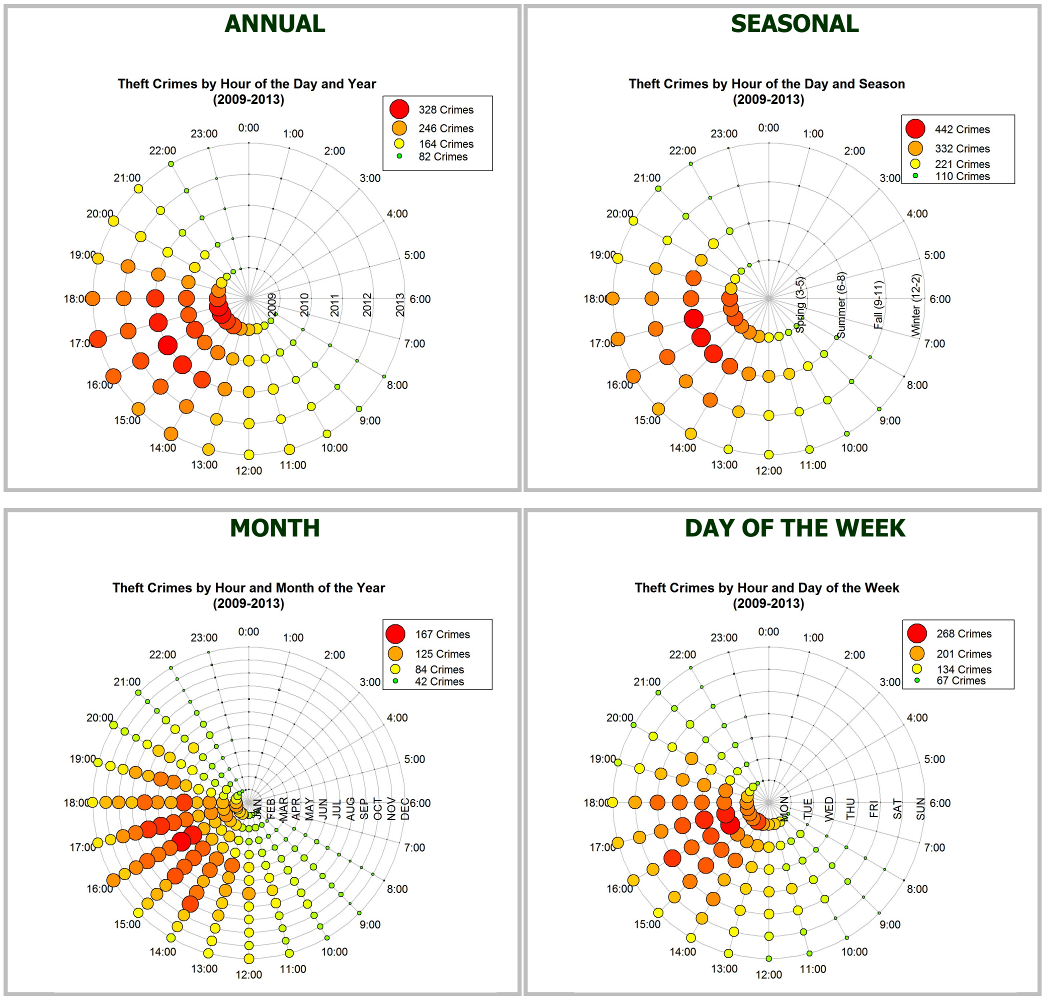
A27. Sexual Assault Crimes Clockplots: Collison Frequency by Year, Season, Month, and Day of the Week.



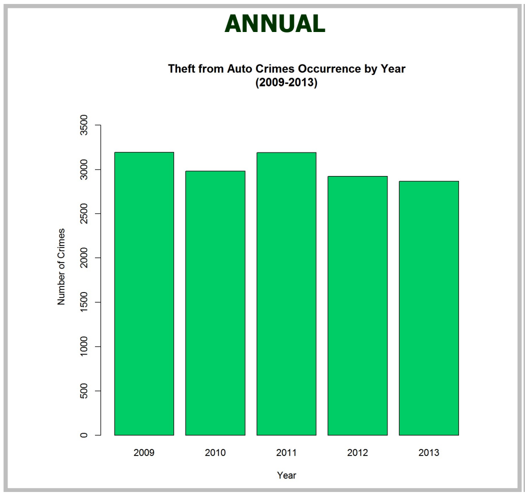
A28. Theft Crimes Bar Charts: Collisions Bar Charts: Collison Frequency by Year, Season, Month, Day of the Week, and Hour of the Day.

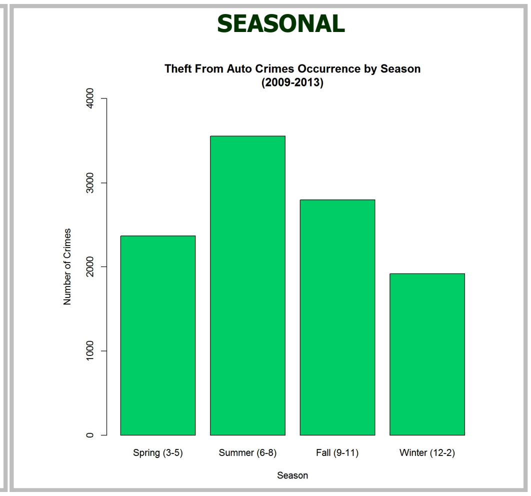


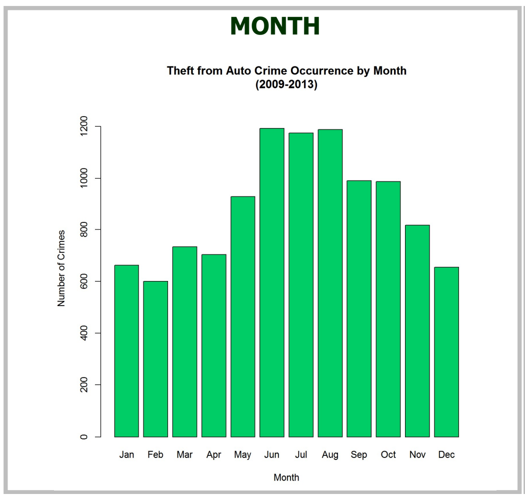
A29. Theft Crimes Clockplots: Collison Frequency by Year, Season, Month, and Day of the Week.

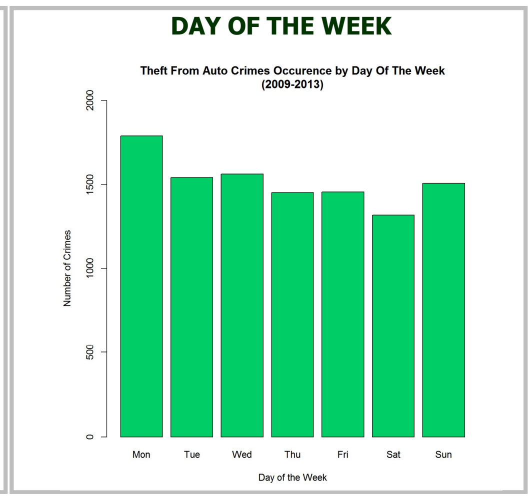


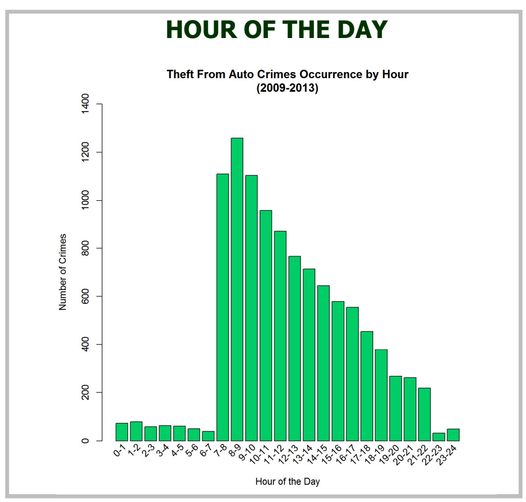
A30. Theft from Auto Crimes Bar Charts: Collisions Bar Charts: Collison Frequency by Year, Season, Month, Day of the Week, and Hour of the Day.



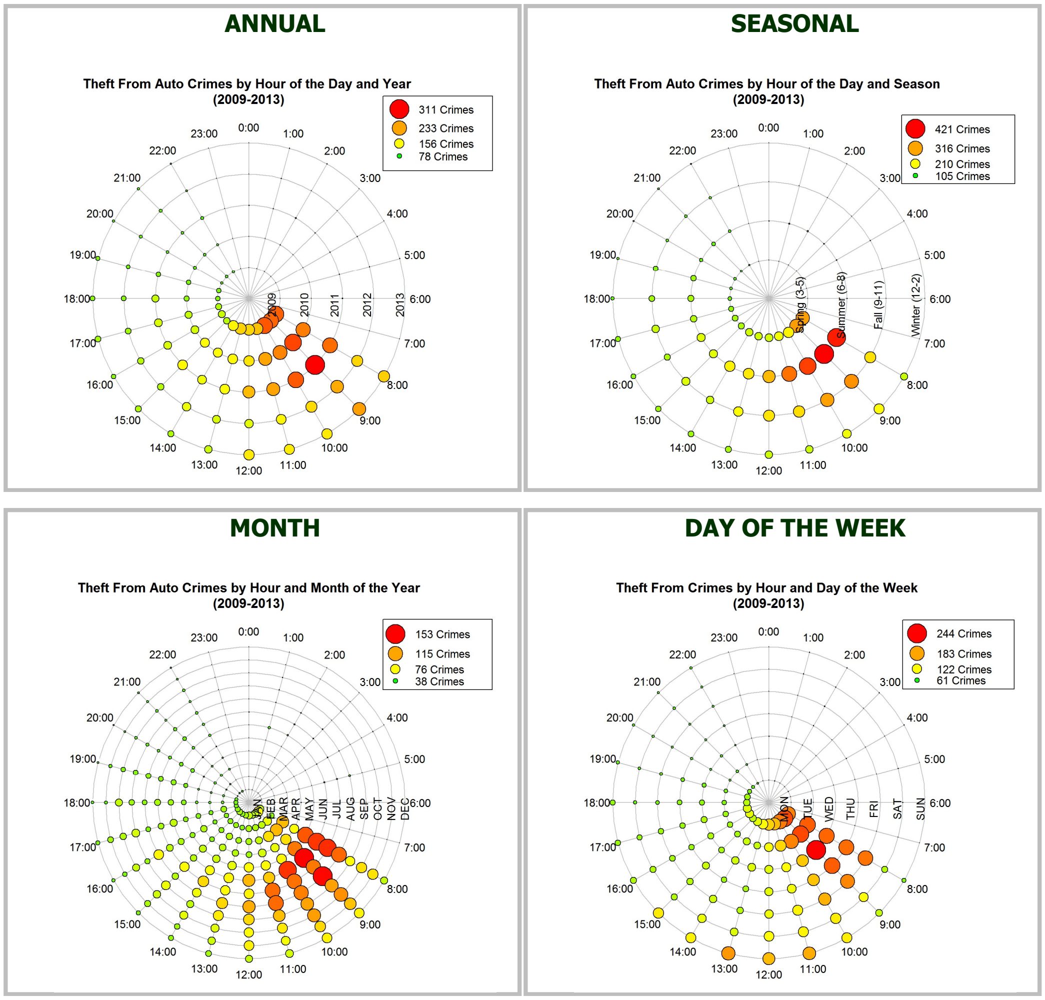




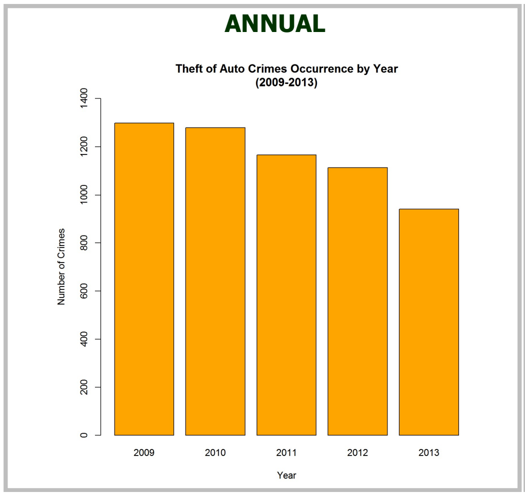


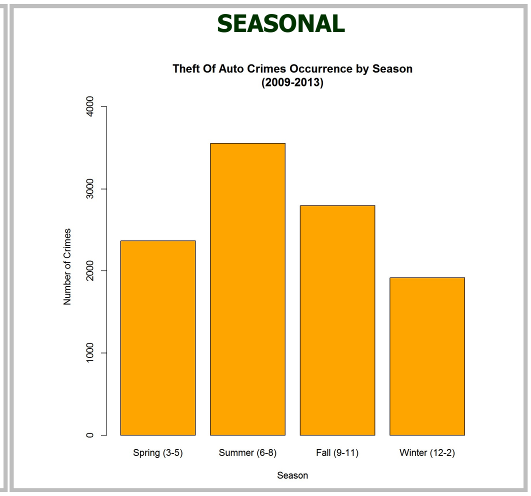


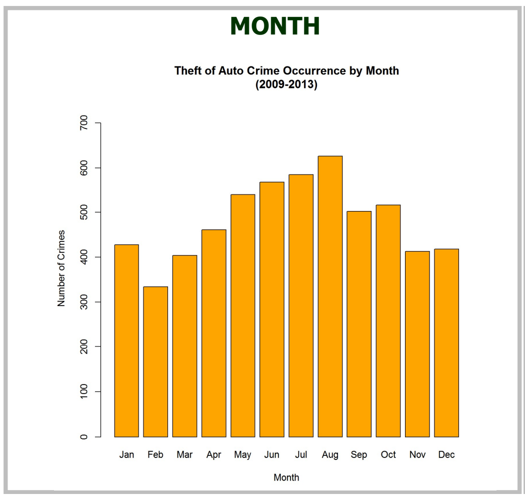
A29. Theft from Auto Crimes Clockplots: Collison Frequency by Year, Season, Month, and Day of the Week.

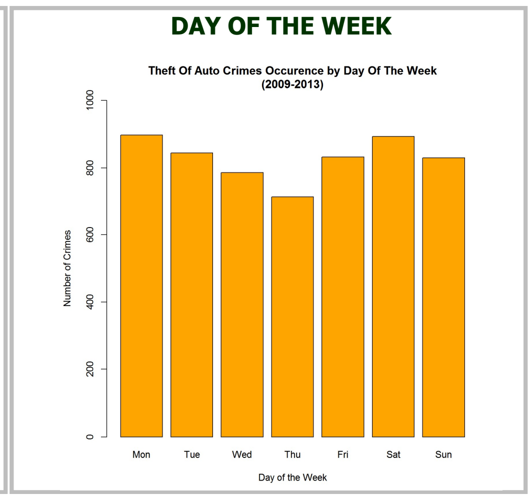


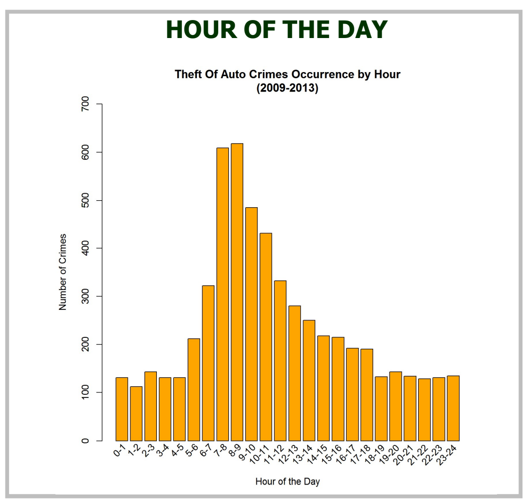
A31. Theft of Auto Crimes Bar Charts: Collisions Bar Charts: Collison Frequency by Year, Season, Month, Day of the Week, and Hour of the Day.



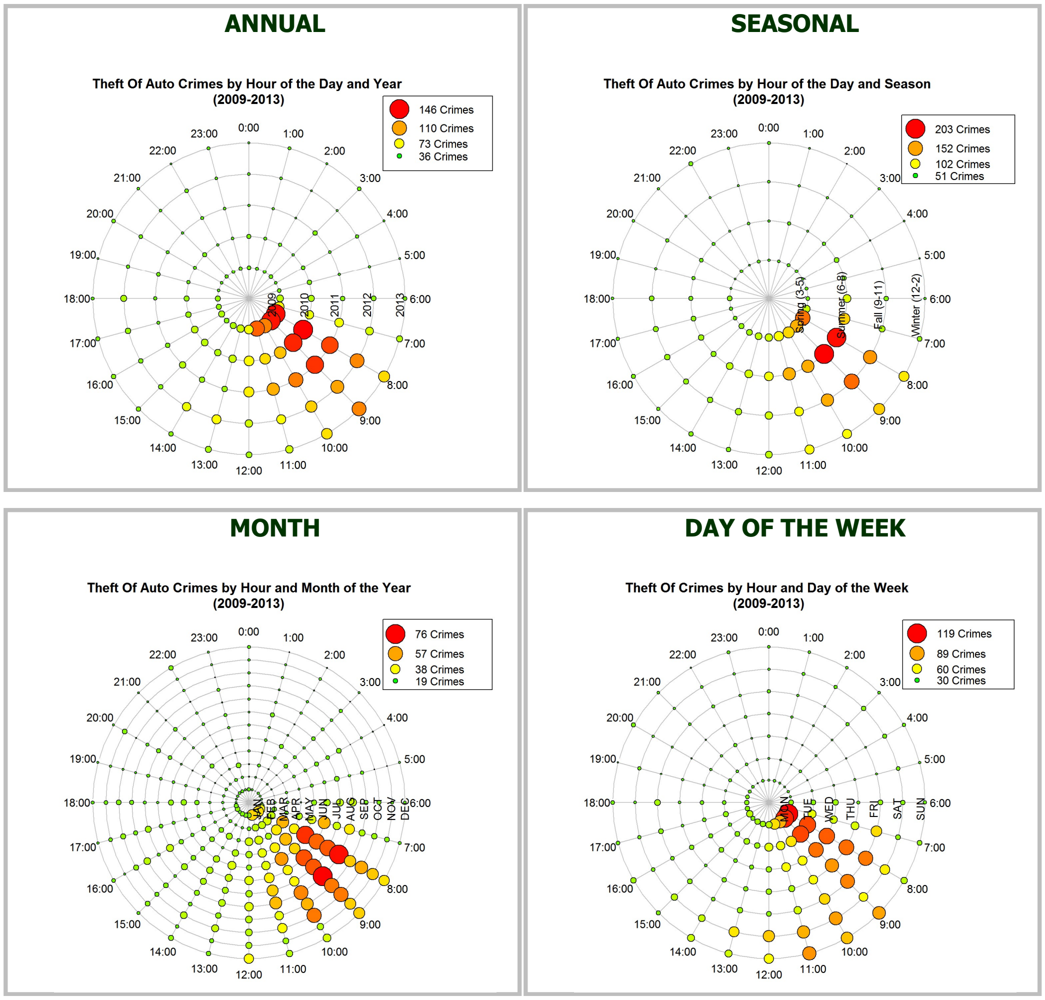








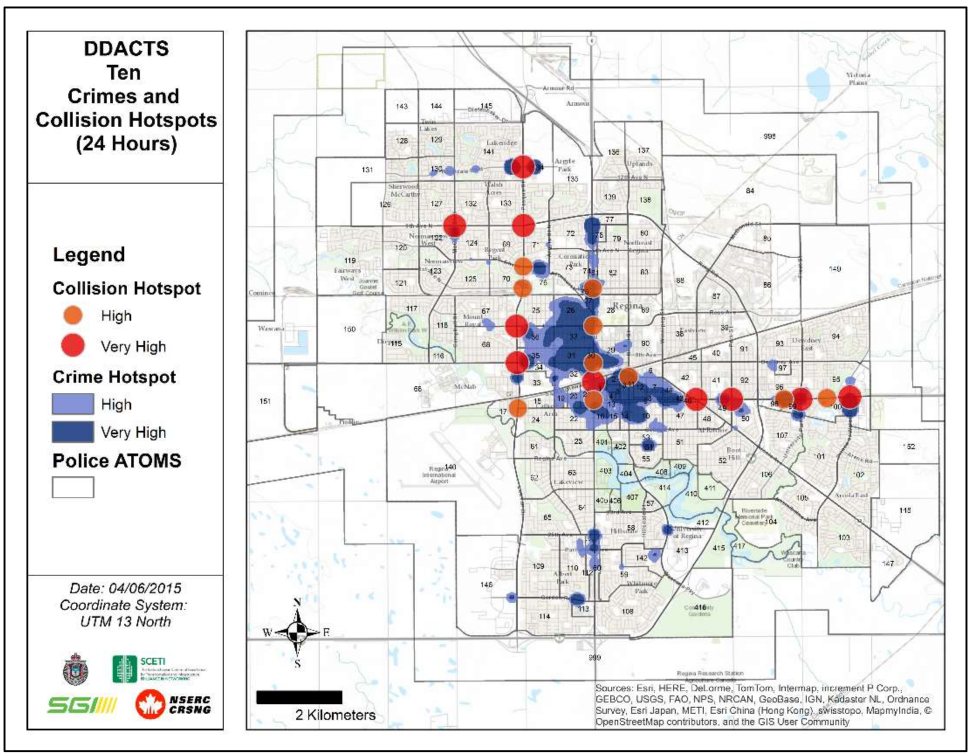
A32. Theft of Auto Crimes Clockplots: Collison Frequency by Year, Season, Month, and Day of the Week.



APPENDIX B: Spatio-Temporal Descriptive Statistics of Data

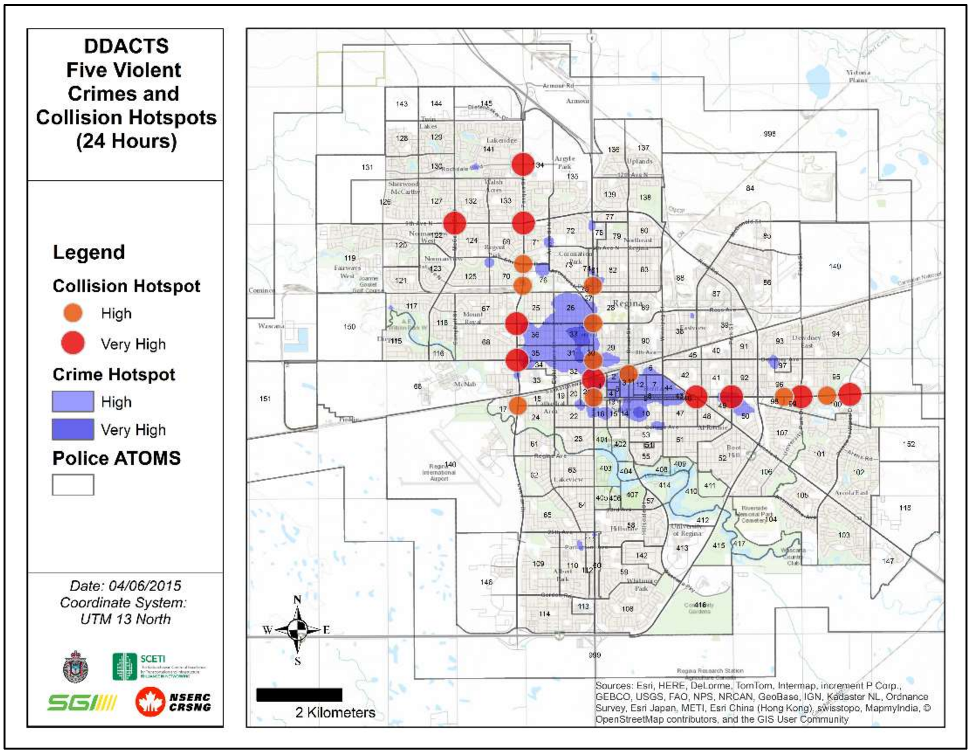
Kernel Density Estimation (KDE) technique was used to analyze and display historical crimes for the City of Regina. Influence areas of 300 squared meters were used in the KDE analysis. Dot technique was also used to display observed traffic collisions; locations with very high frequencies of collisions are displayed with red dots with bigger radius and locations with high collision frequencies are displayed with orange dots. Locations with collision frequencies were ranked and the top 10 ranked locations were assigned very and top 11 to 20 were designated as high. Tables representing these collision frequencies by year will be presented in this appendix.

B1: 24 Hours Total Crimes and Total Collisions Hotspots

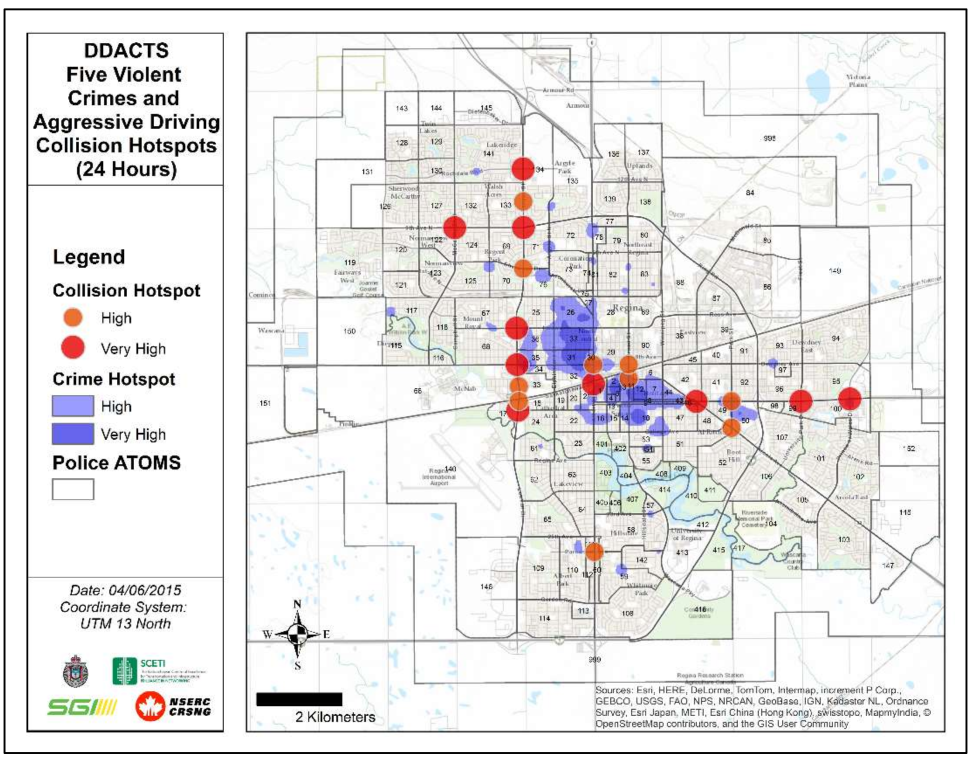


B2: Violent Crimes and Collisions Hotspots

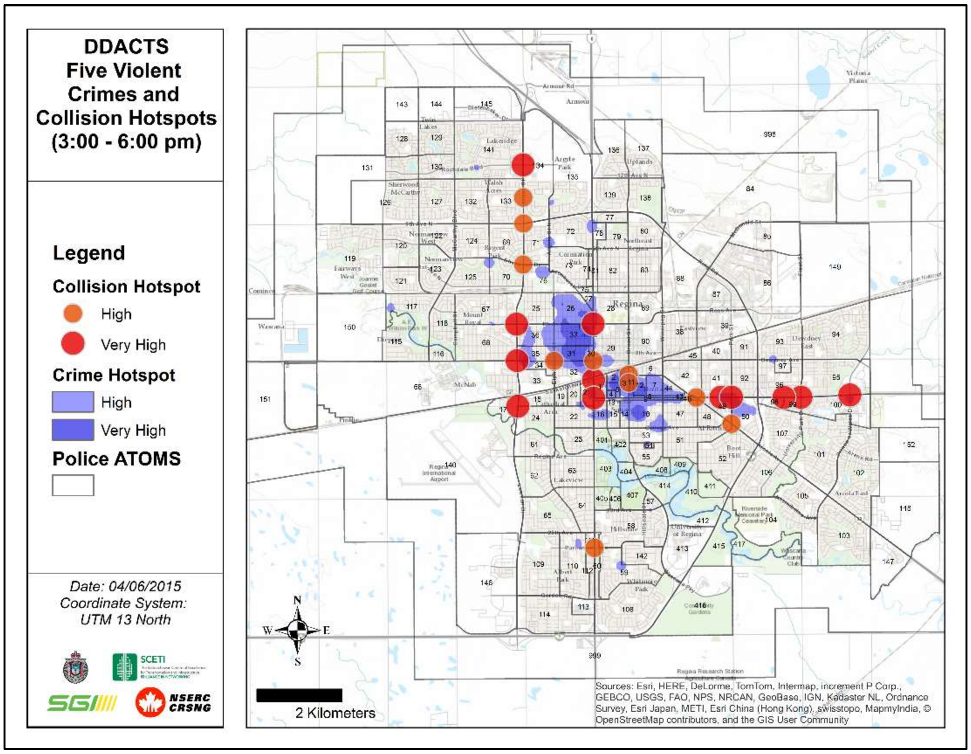
24 Hours Violent Crimes and Total Collisions Hotspots



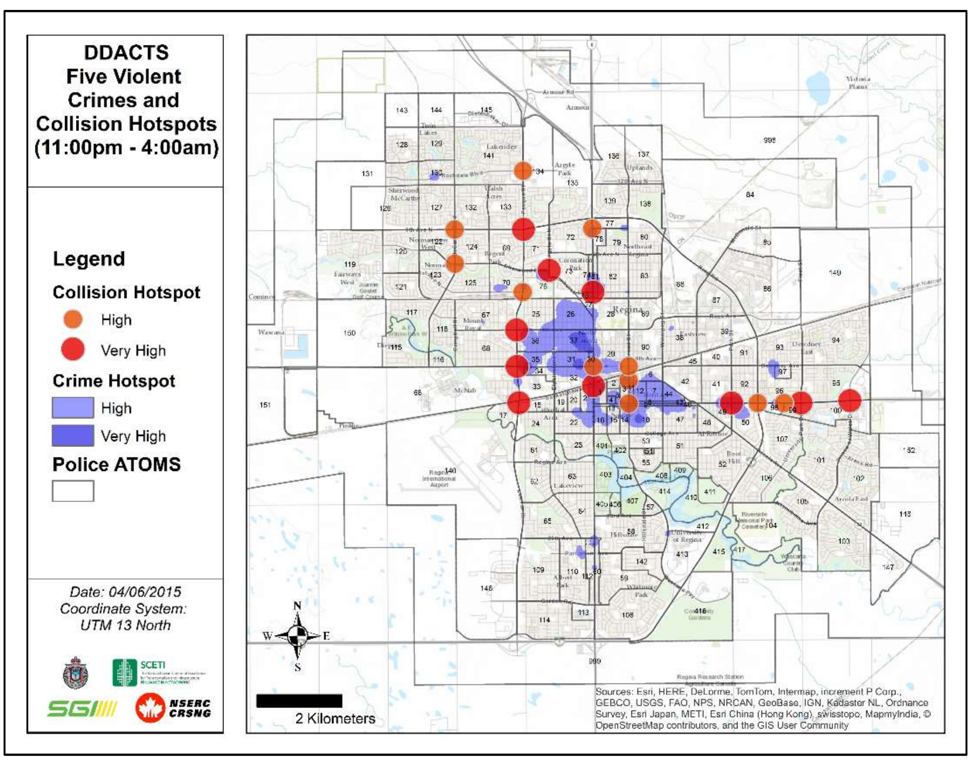
24 Hours Violent Crimes and Aggressive Driving Collisions Hotspots



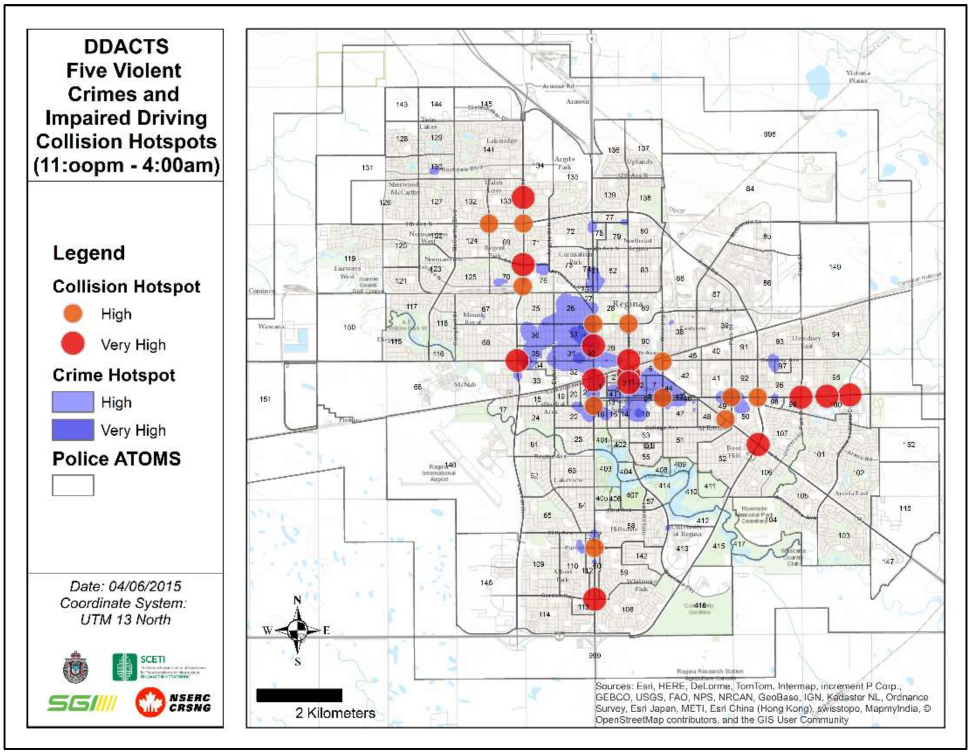
3:00 – 6:00 pm Peak Hour Violent Crimes and Total Collisions Hotspots



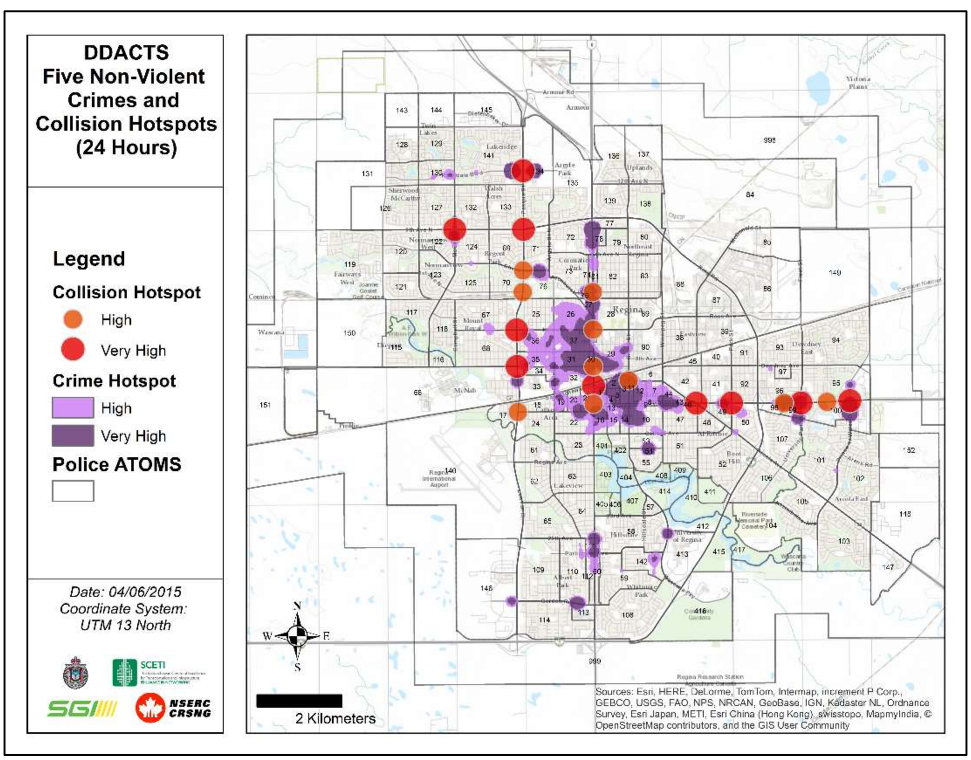
11:00 pm– 4:00 am Peak Hour Violent Crimes and Total Collisions Hotspots



11:00 pm– 4:00 am Peak Hour Violent Crimes and Impaired Driving Collisions Hotspots

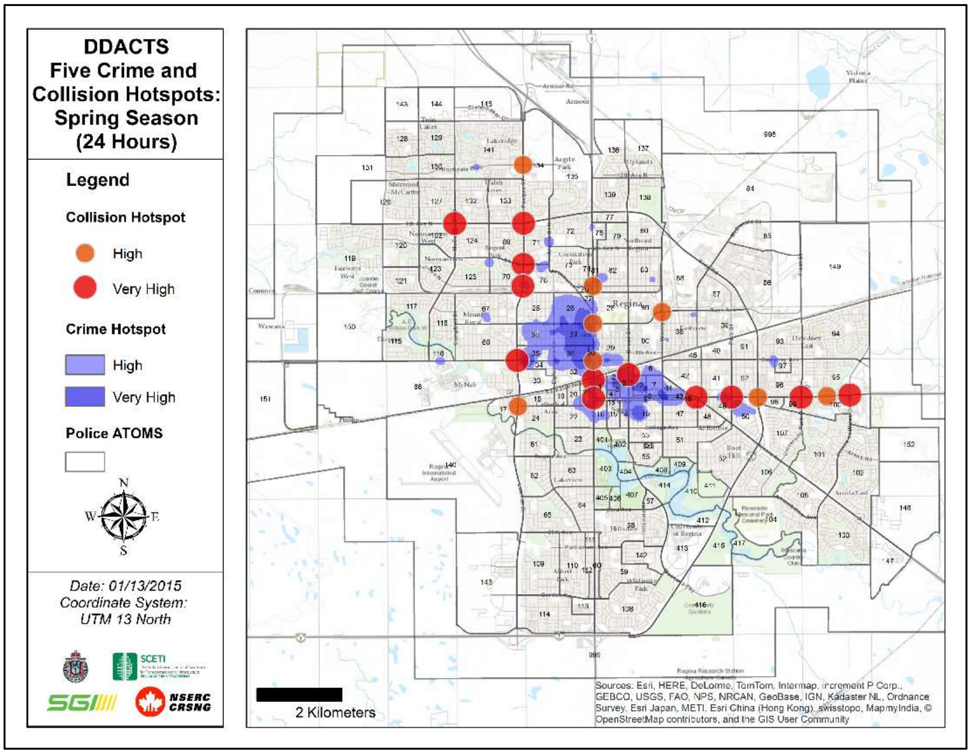


B3: 24 Hours Non-Violent Crimes and Total Collisions Hotspots

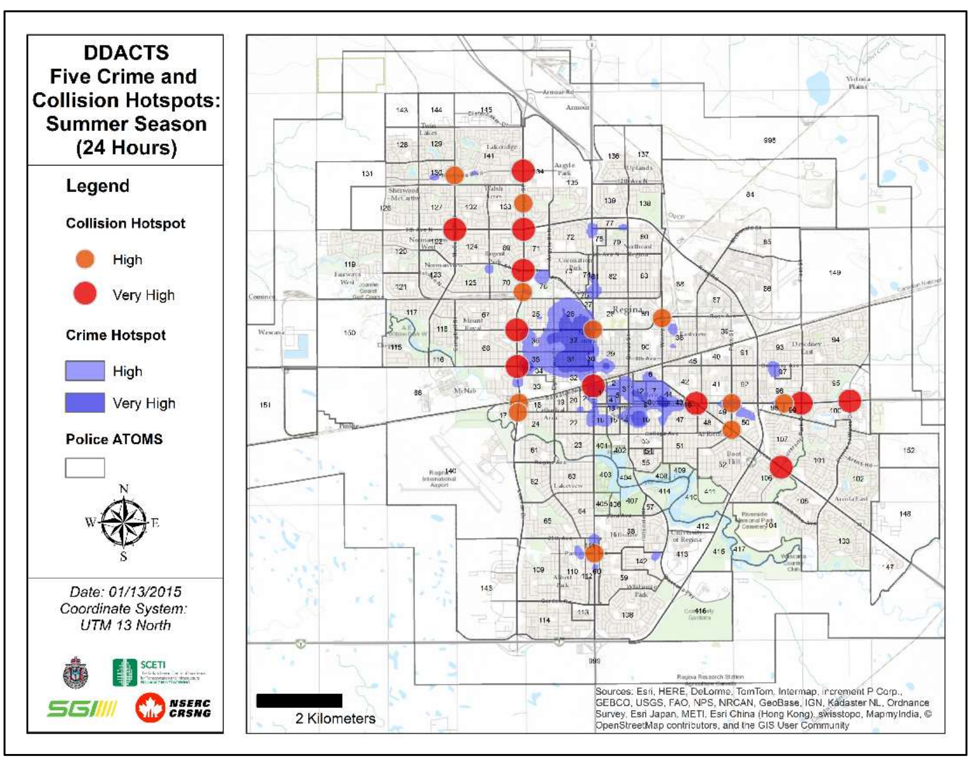


B4: Seasonal Trend: Violent Crimes and Collisions Hotspots

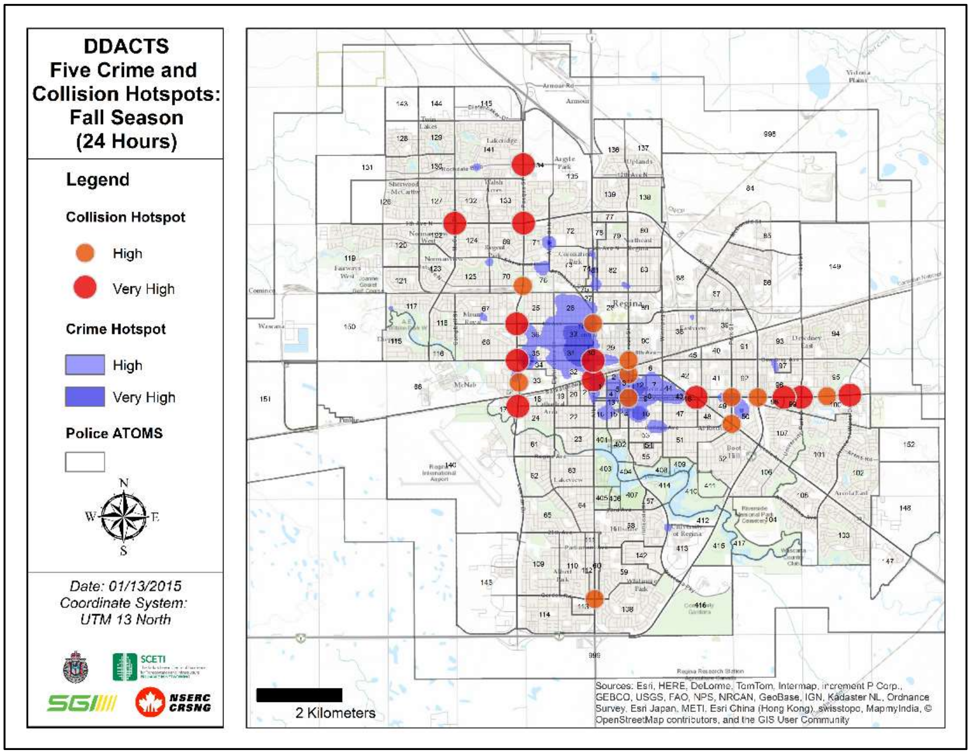
Spring (March, April, and May) Violent Crimes and Total Collisions Hotspots



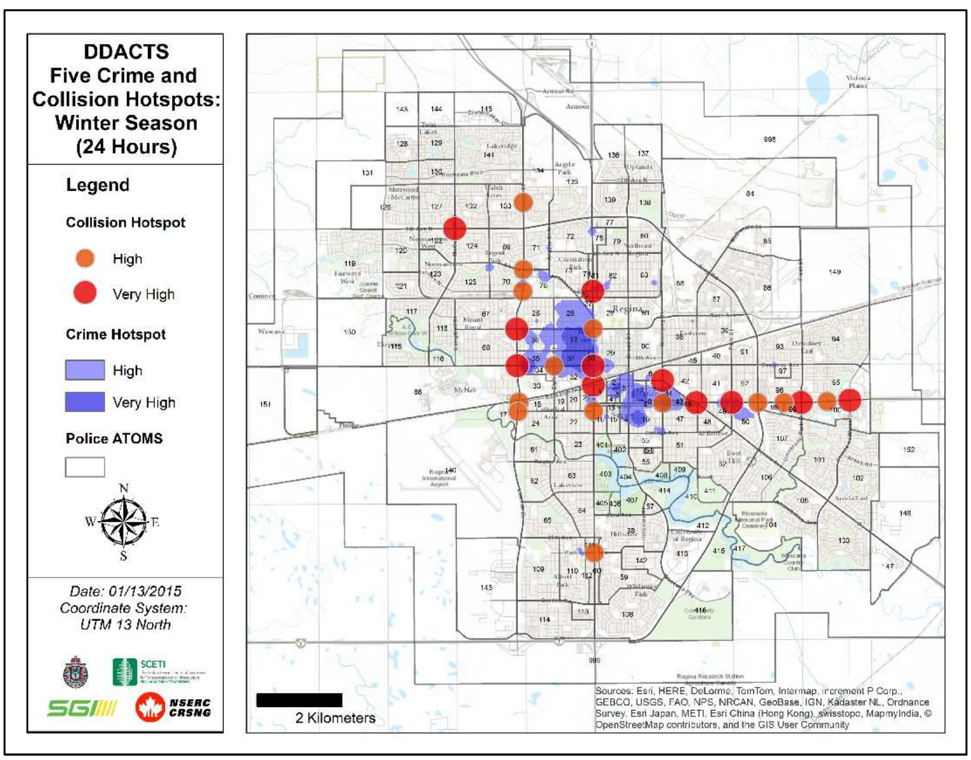
Summer (June, July and August) Violent Crimes and Total Collisions Hotspots



Fall (September, October, and November) Violent Crimes and Total Collisions Hotspots

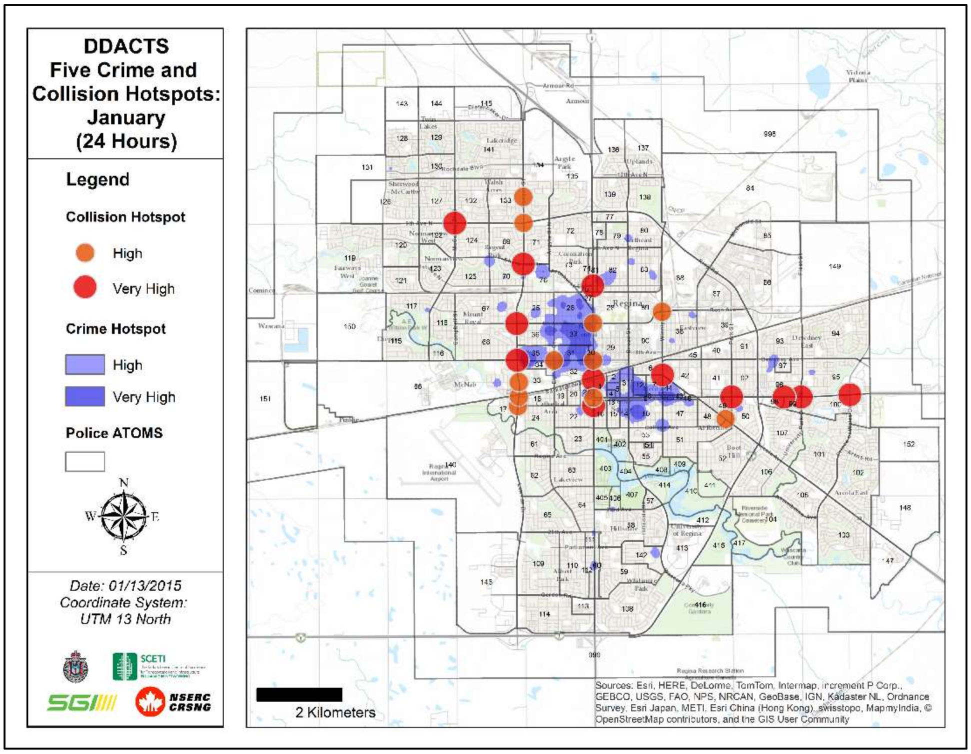


Winter (December, January, and February) Violent Crimes and Total Collisions Hotspots

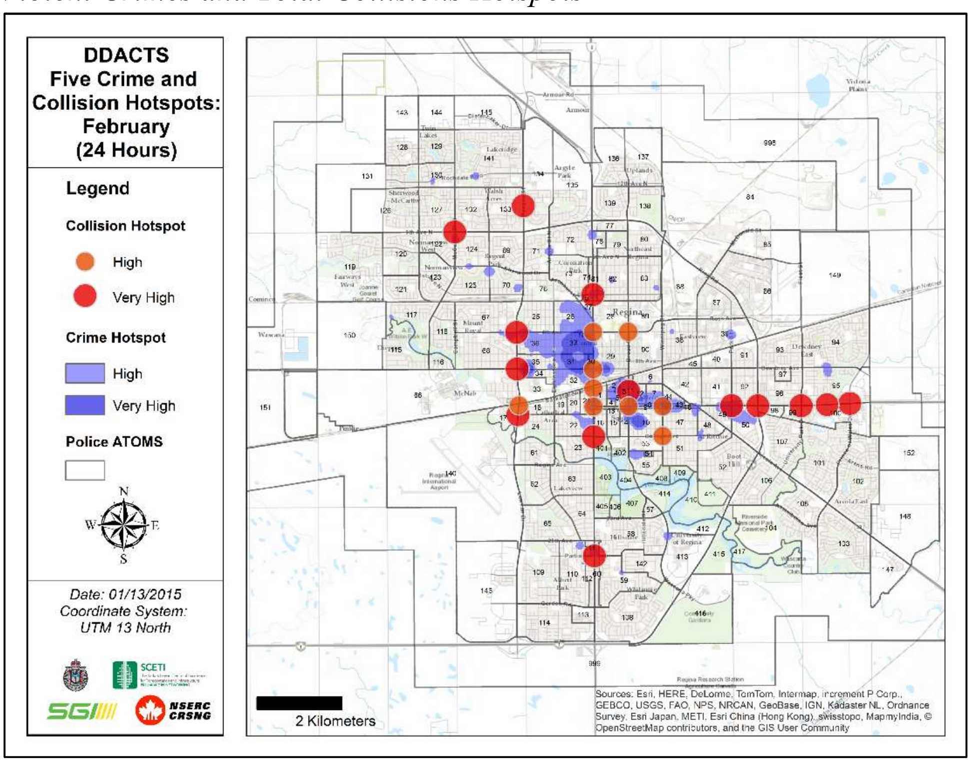


B5: Monthly Trend: Violent Crimes and Collisions Hotspots

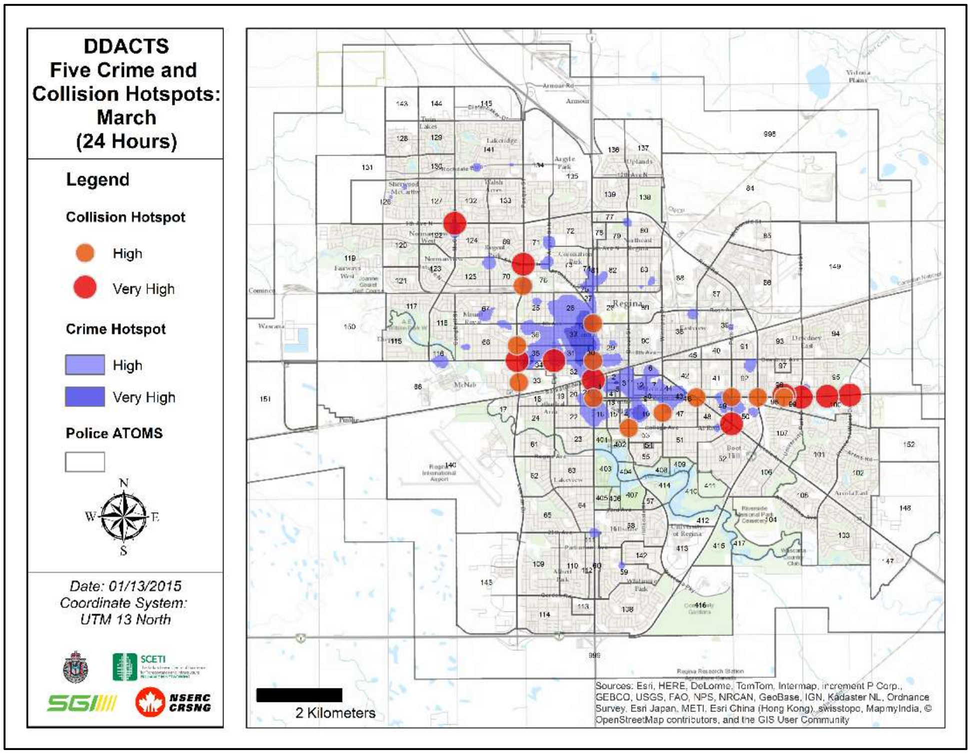
January: Violent Crimes and Total Collisions Hotspots



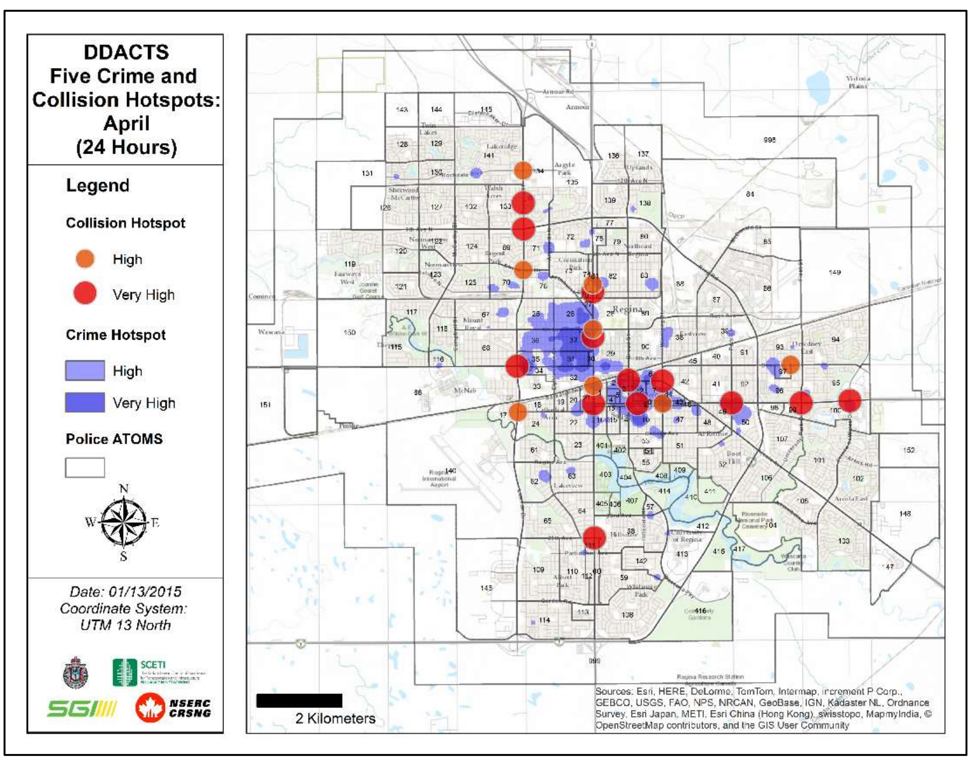
February: Violent Crimes and Total Collisions Hotspots



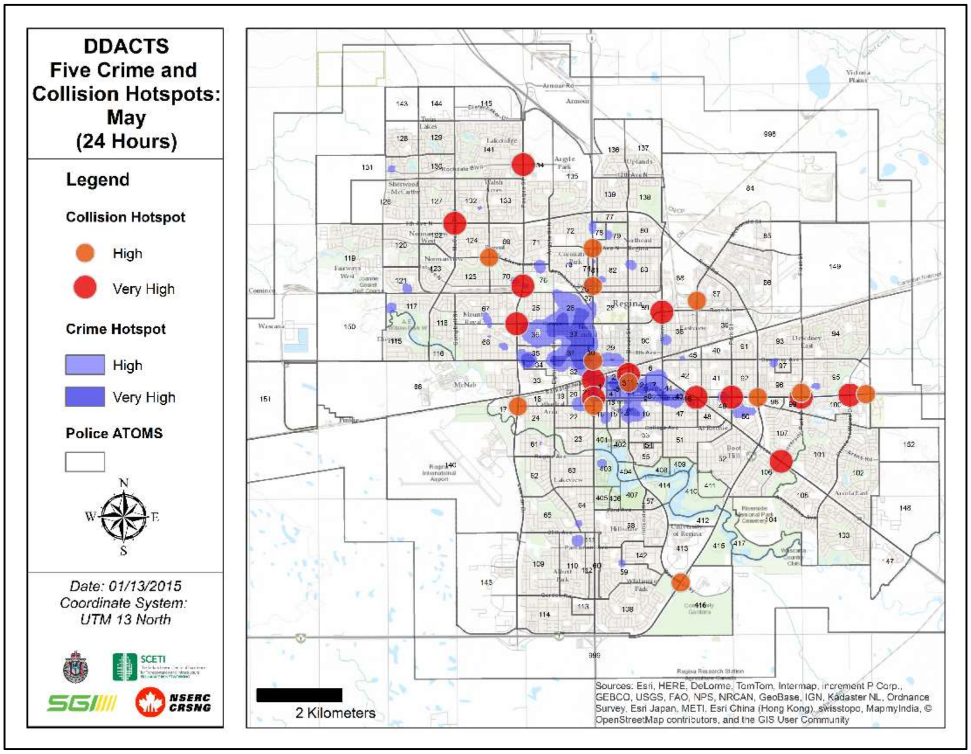
March: Violent Crimes and Total Collisions Hotspots



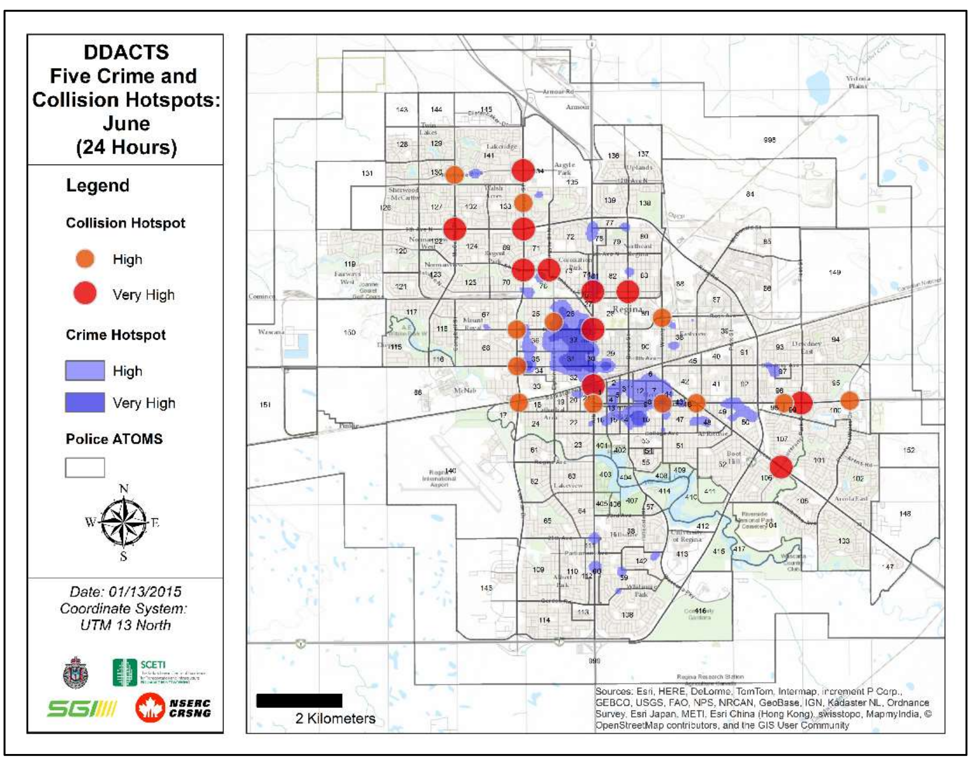
April: Violent Crimes and Total Collisions Hotspots



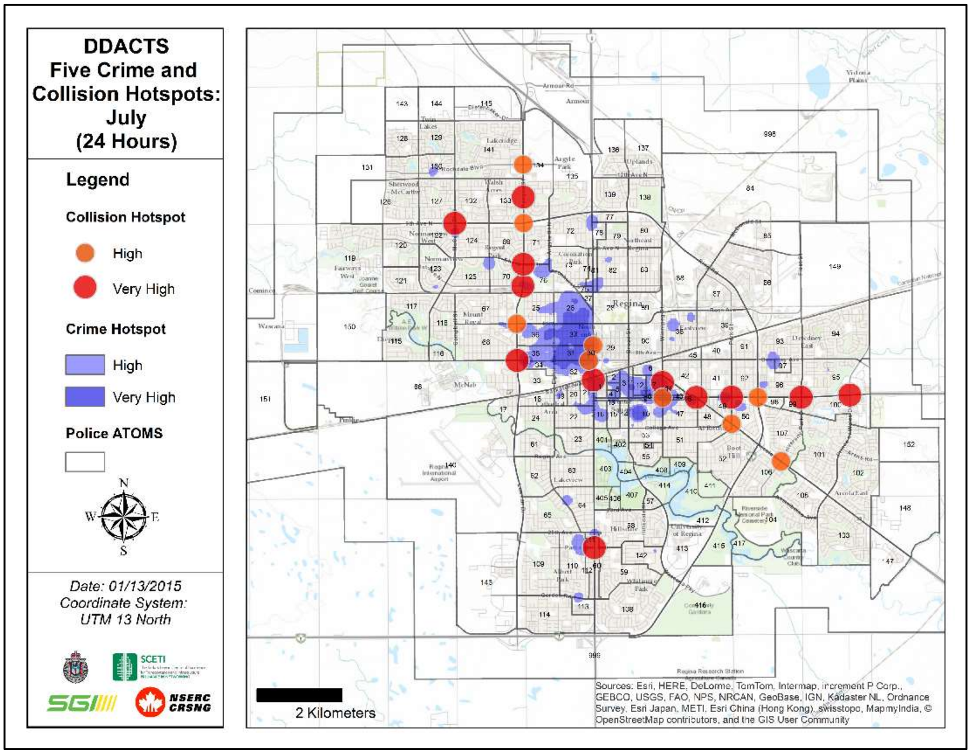
May: Violent Crimes and Total Collisions Hotspots



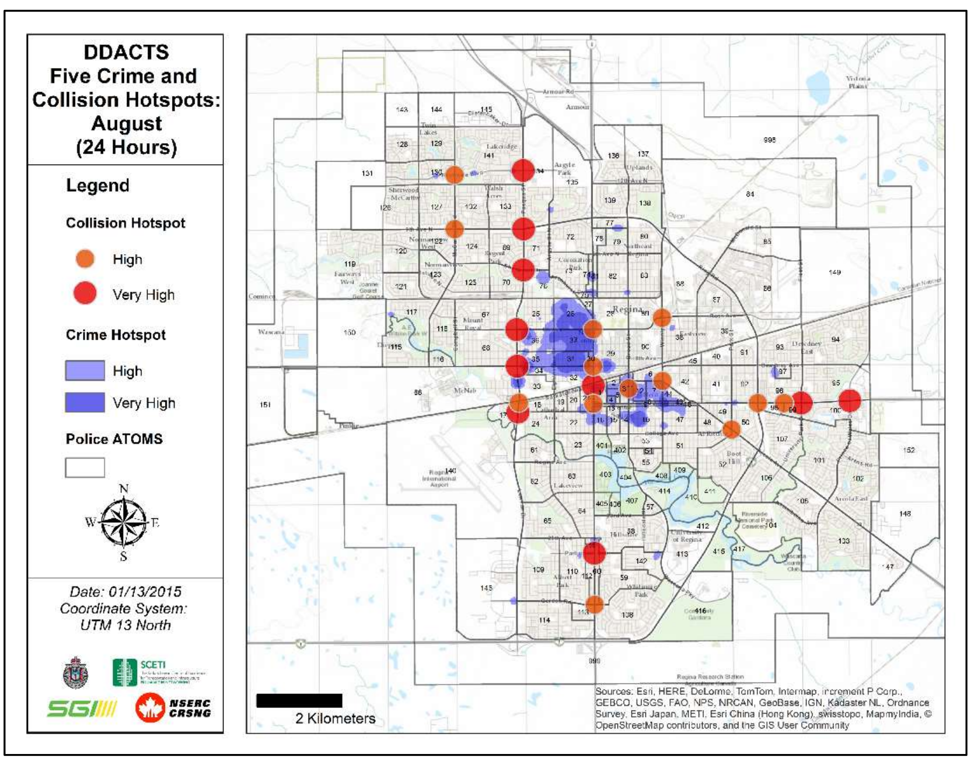
June: Violent Crimes and Total Collisions Hotspots



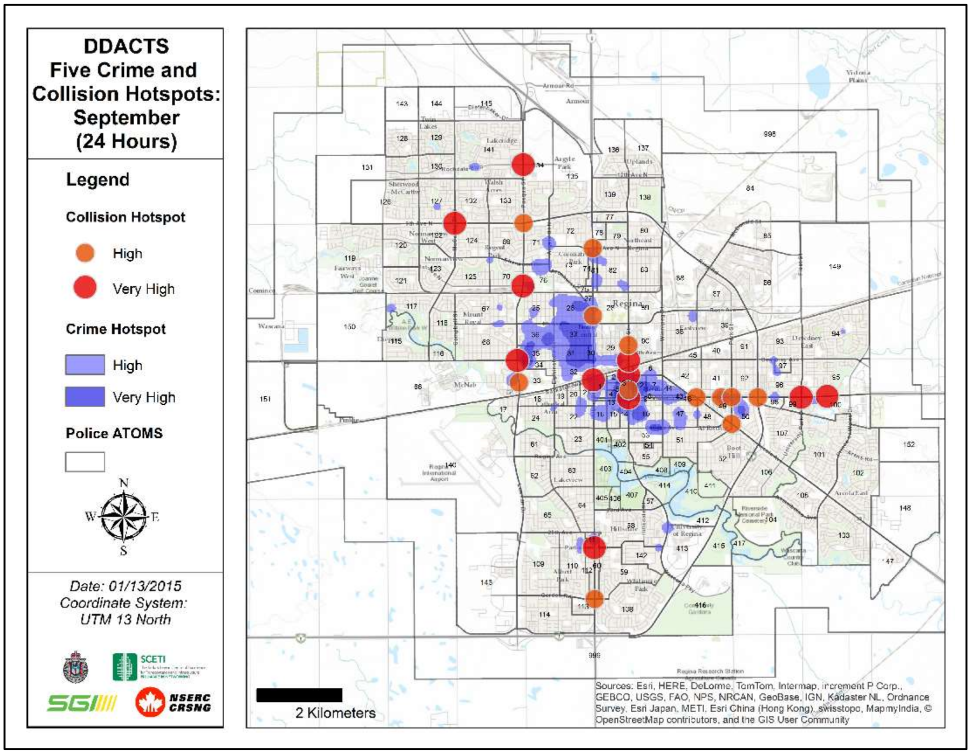
July: Violent Crimes and Total Collisions Hotspots



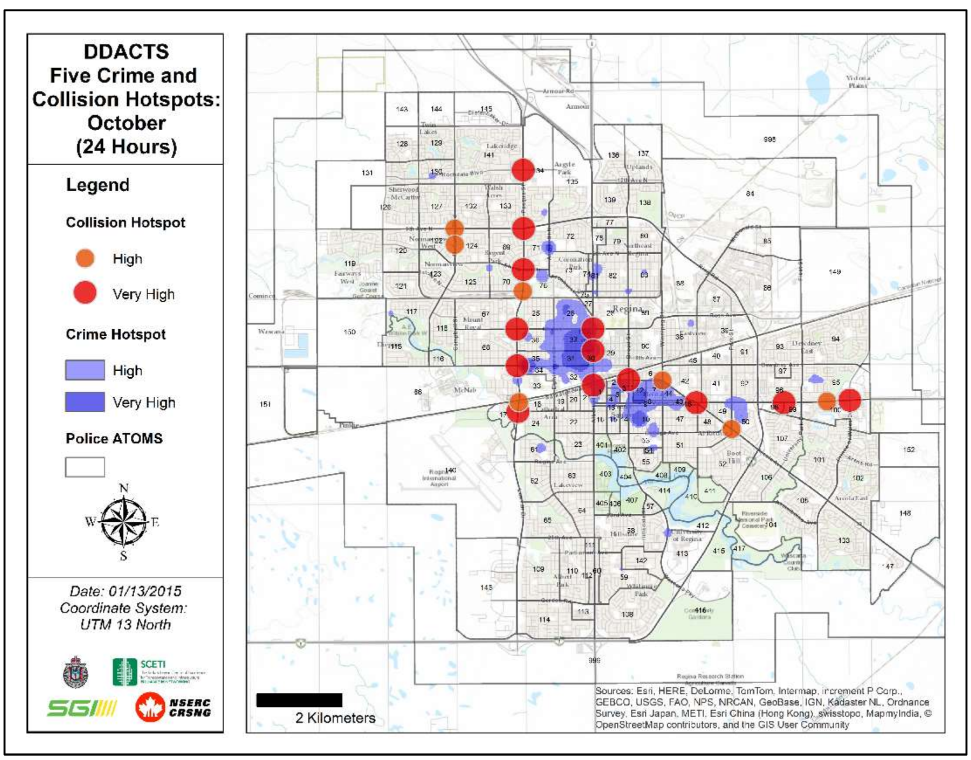
August: Violent Crimes and Total Collisions Hotspots



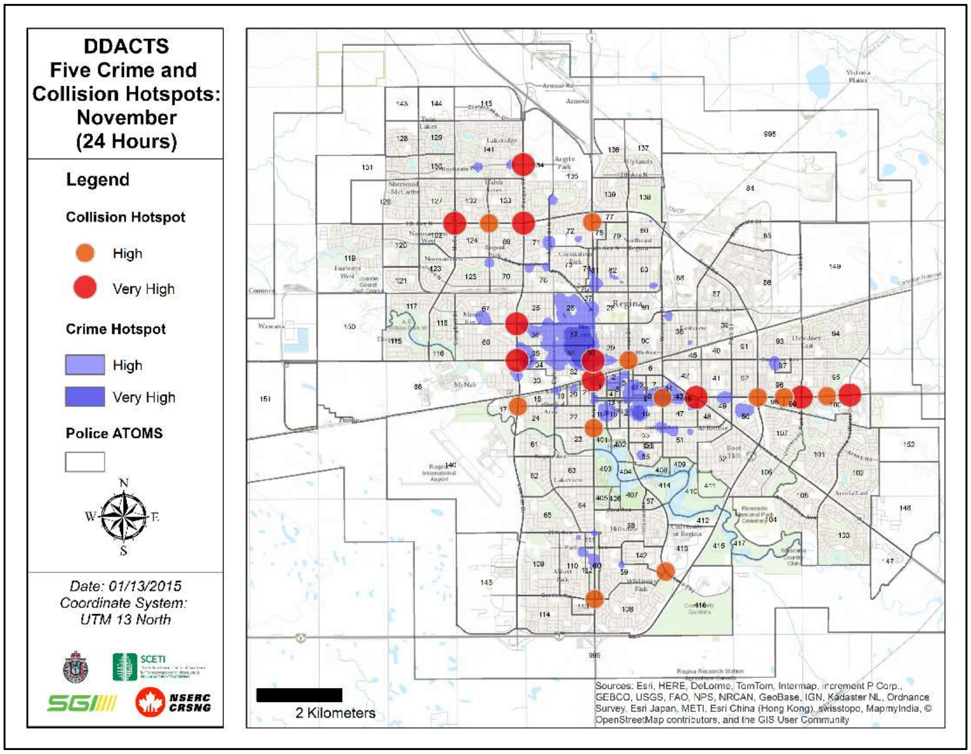
September: Violent Crimes and Total Collisions Hotspots



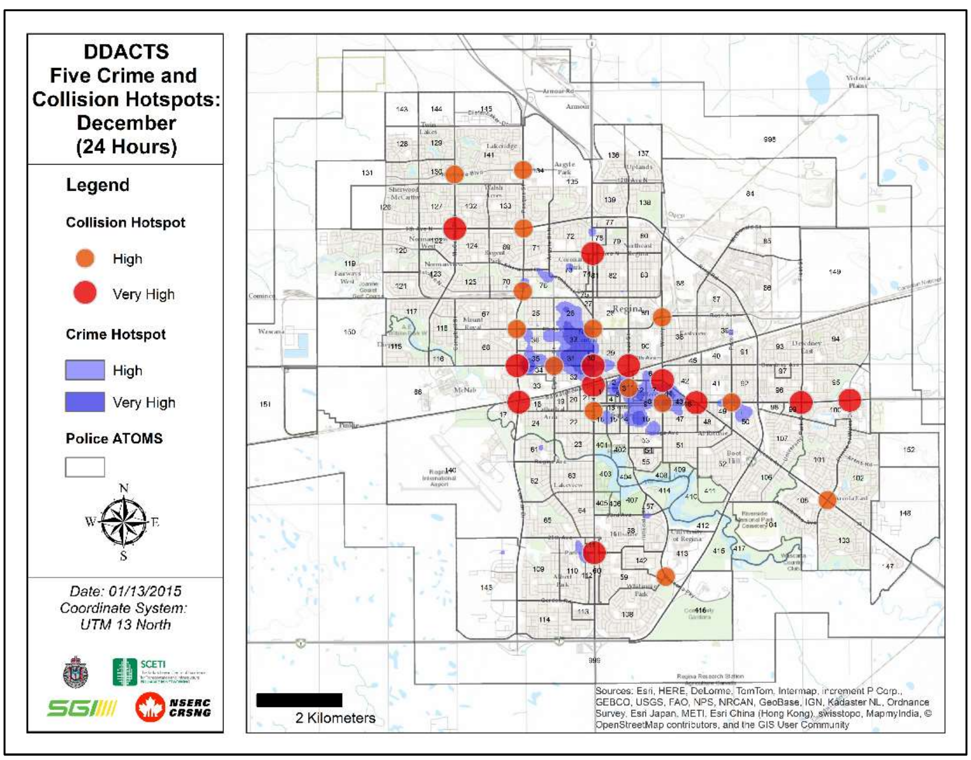
October: Violent Crimes and Total Collisions Hotspots



November: Violent Crimes and Total Collisions Hotspots

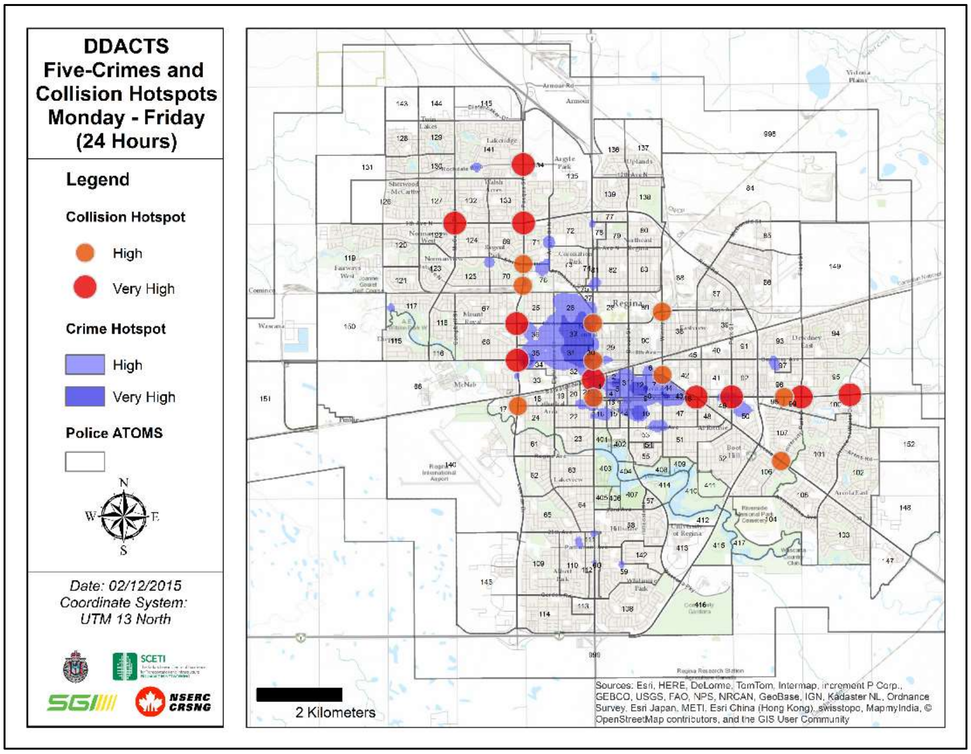


December: Violent Crimes and Total Collisions Hotspots

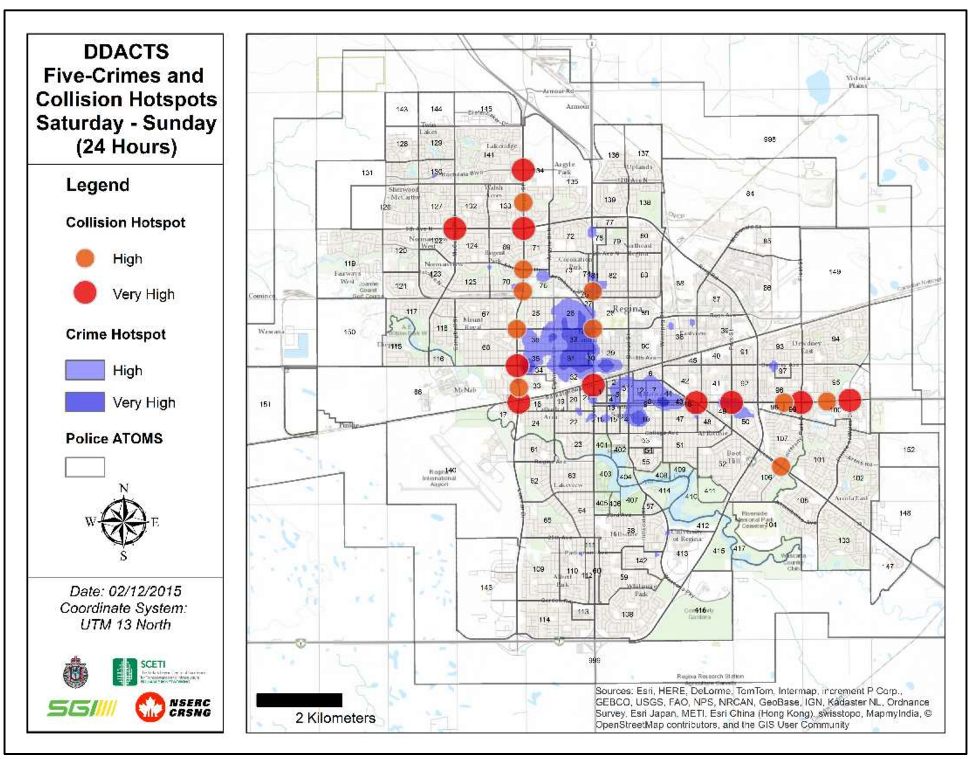


B6: Daily Trend: Violent Crimes and Collisions Hotspots

Monday – Friday: Violent Crimes and Total Collisions Hotspots



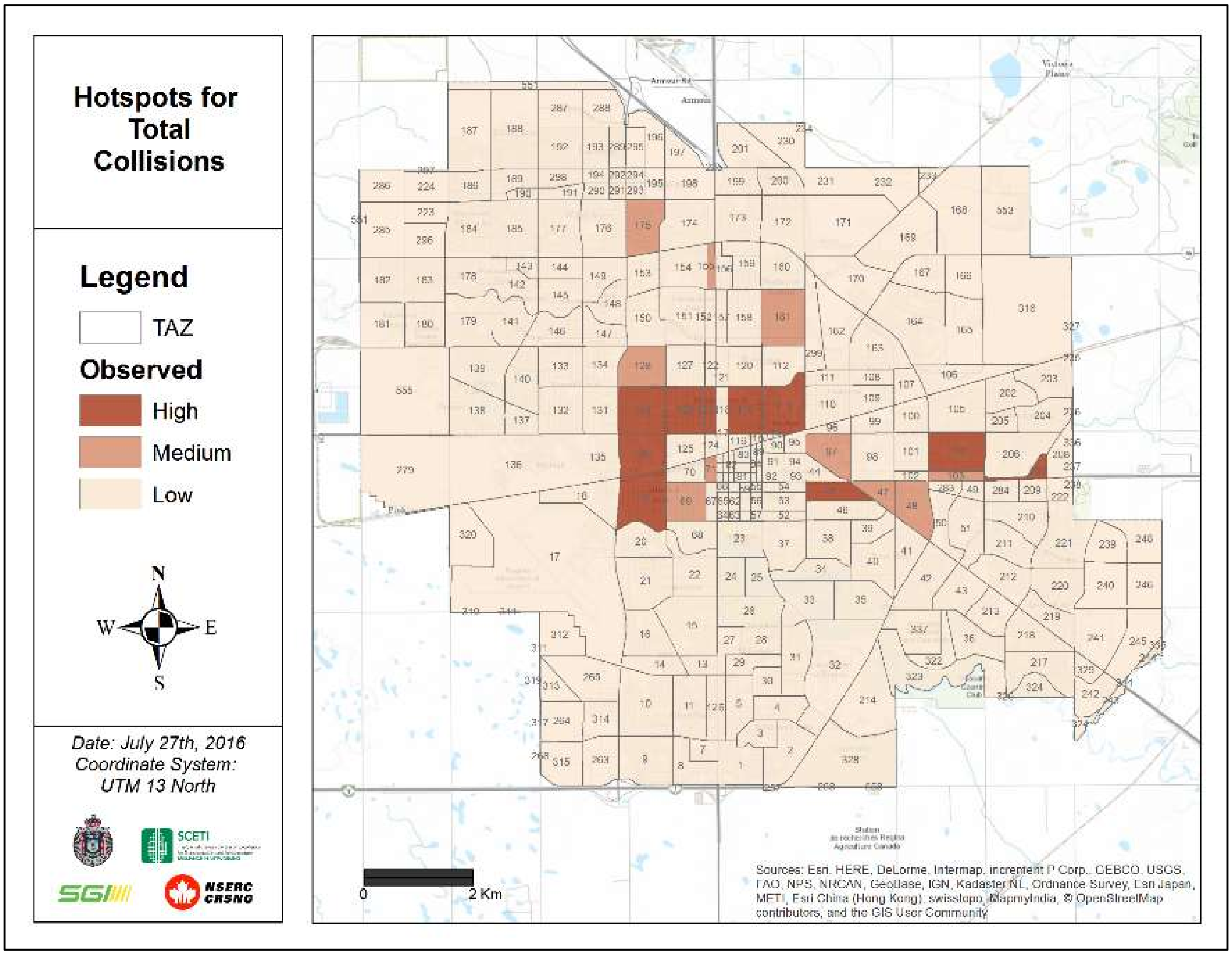
Saturday and Sunday: Violent Crimes and Total Collisions Hotspots



APPENDIX C: Spatial Descriptive Statistics of Data

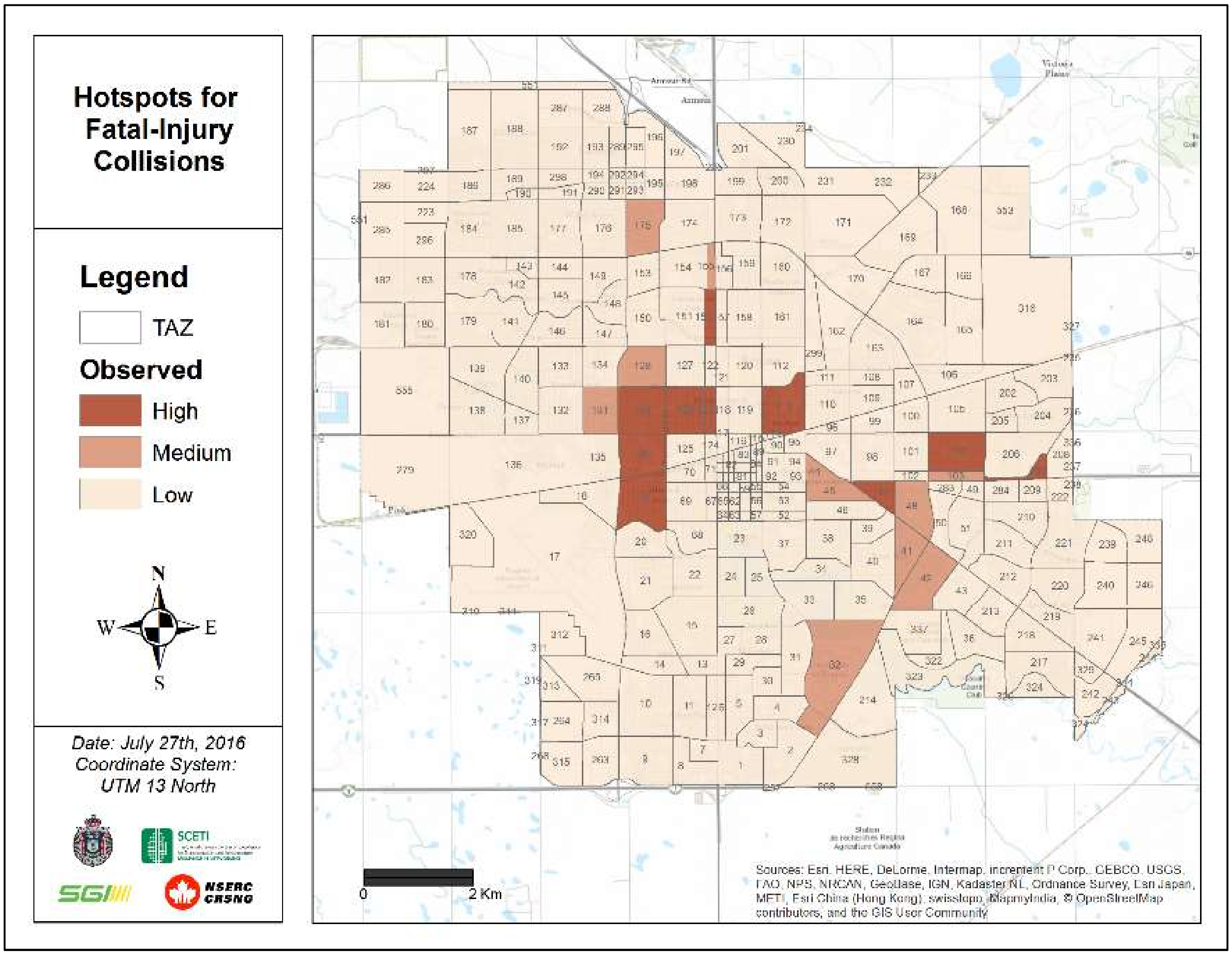
C1. Traffic Collisions

Observed Total Collisions



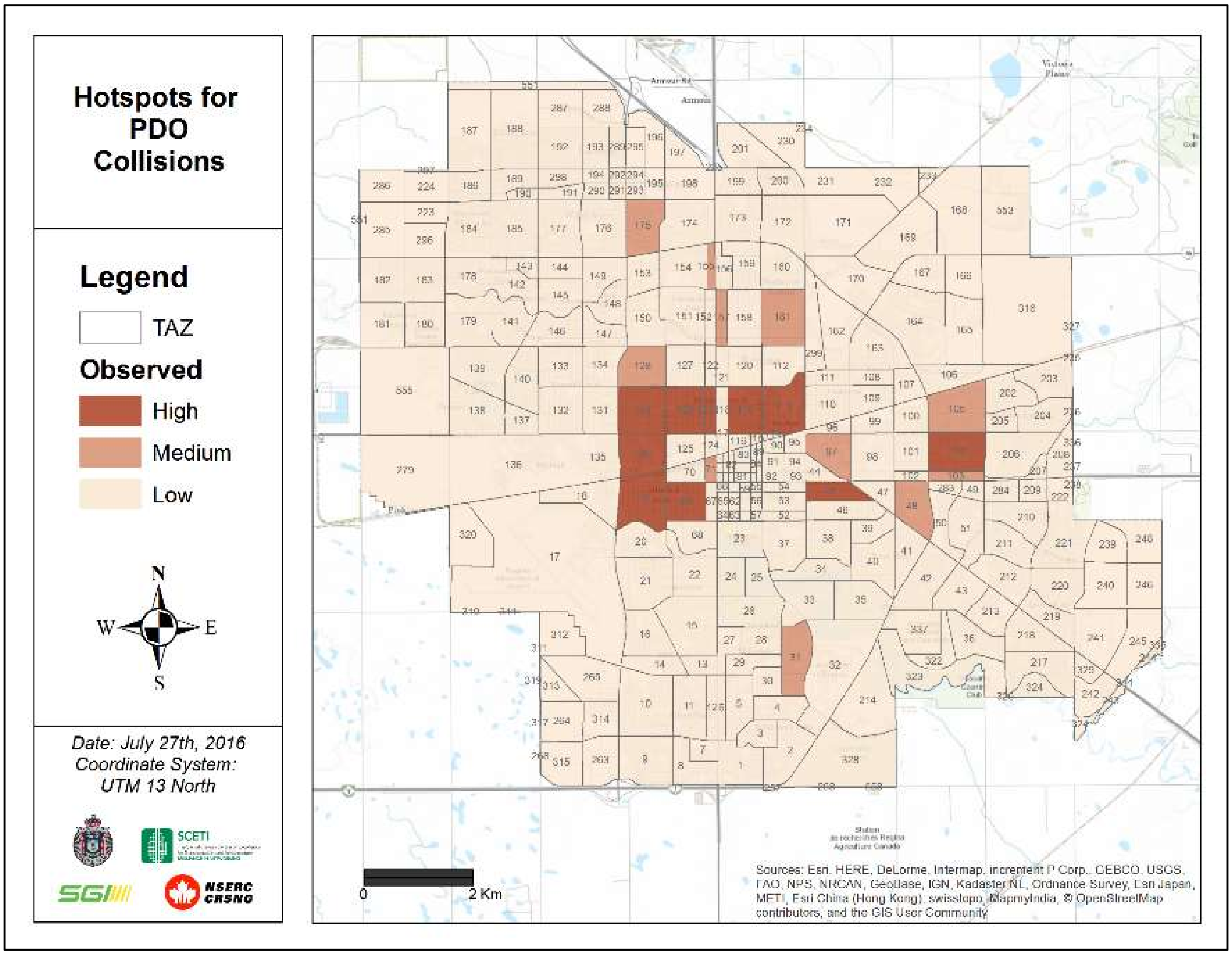
Variable	Observations	Total	Minimum	Maximum	Mean	Standard Deviation
Total Collisions	263	26610	0	424	101.56	88.72

Observed Fatal-Injury Collisions



Variable	Observations	Total	Minimum	Maximum	Mean	Standard Deviation
FI Collisions	263	5759	0	103	21.98	21.65

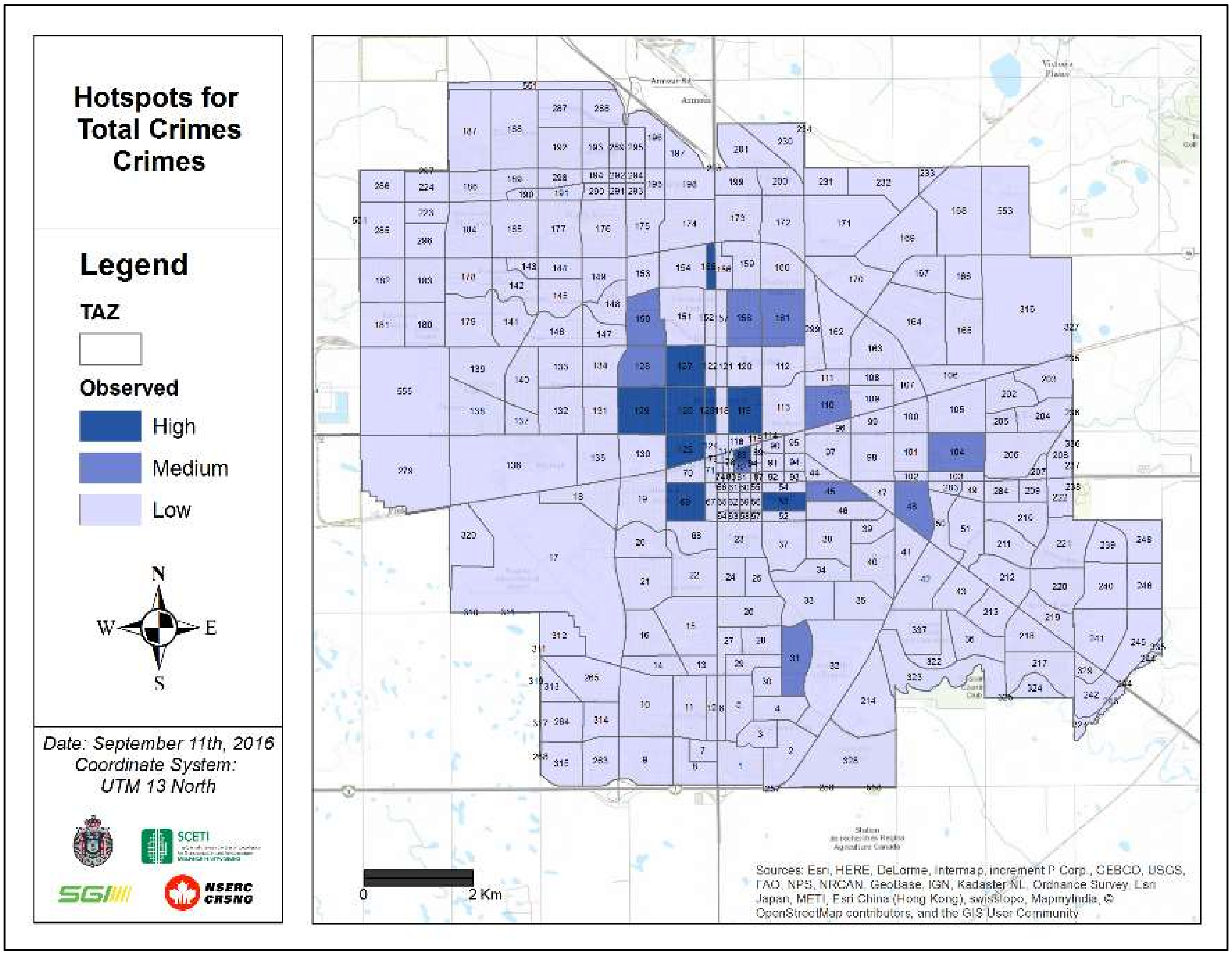
Observed Property Damaged Only Collisions



Variable	Observations	Total	Minimum	Maximum	Mean	Standard Deviation
Property Damage Only Collisions	263	20883	0	326	79.71	69.04

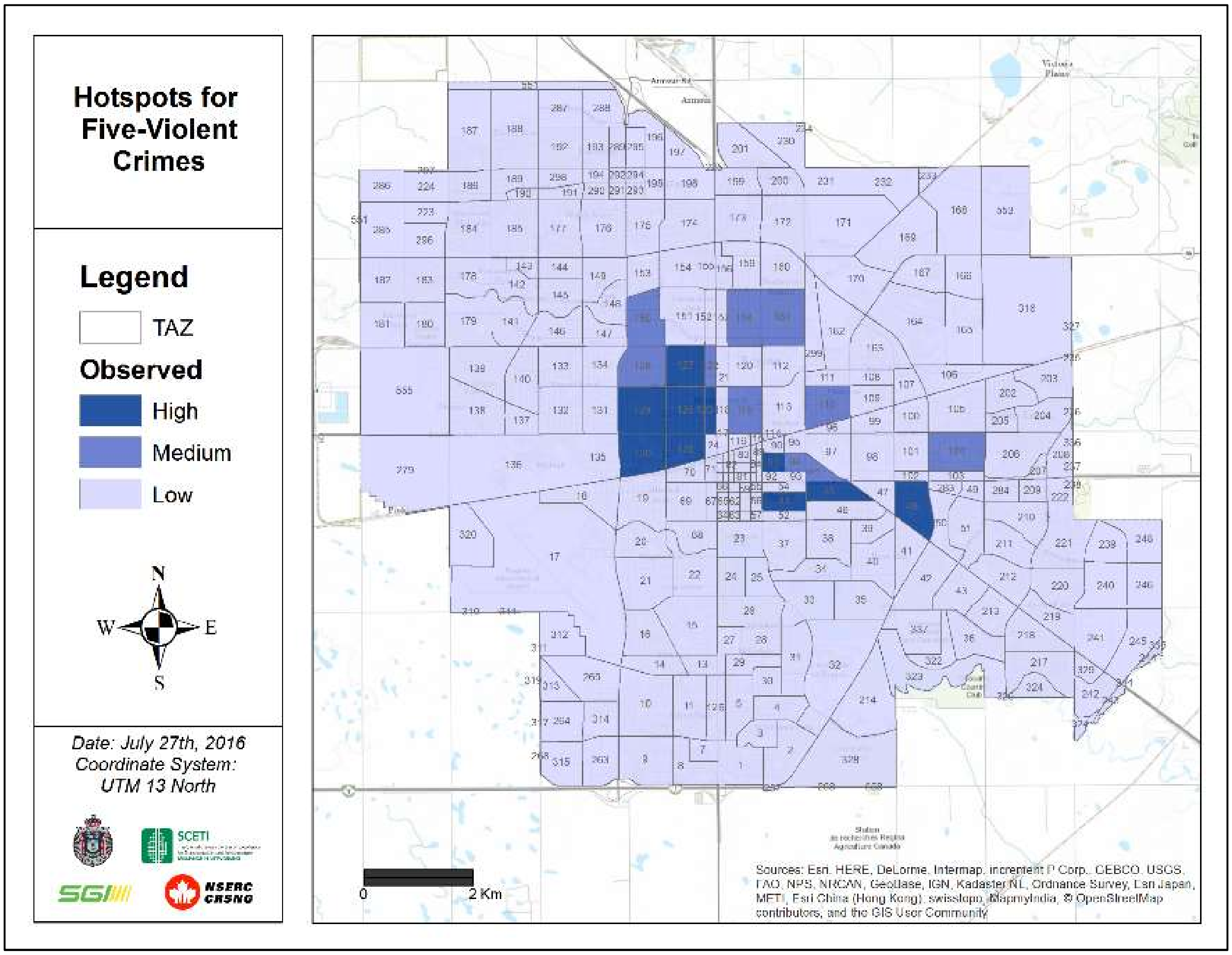
C2. Crimes

Observed Total Crimes



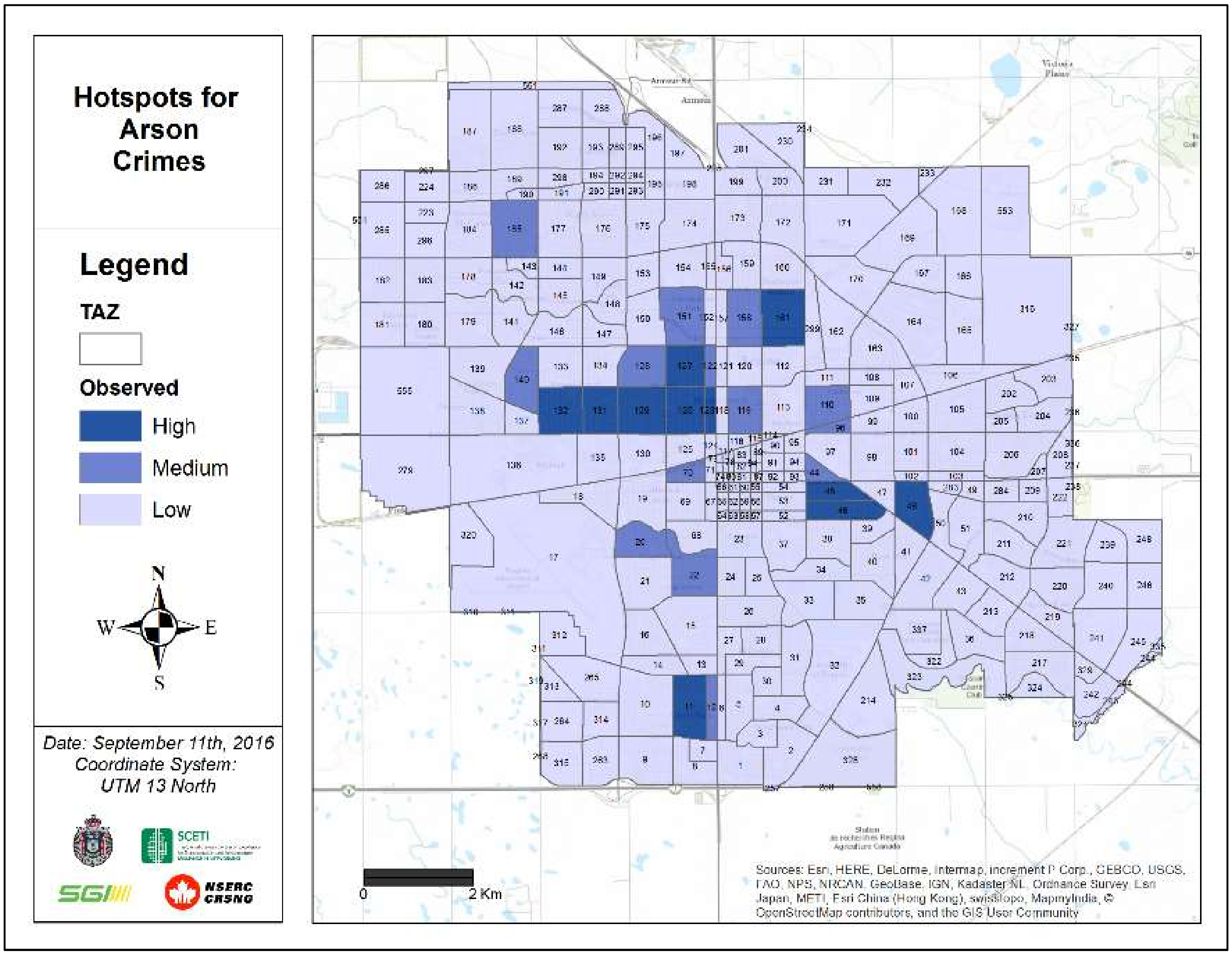
Variable	Observations	Total	Min.	Max.	Mean	Standard Deviation
Total Crimes	263	50148.5	0	2422	173.52	243.25

Observed Violent Crimes



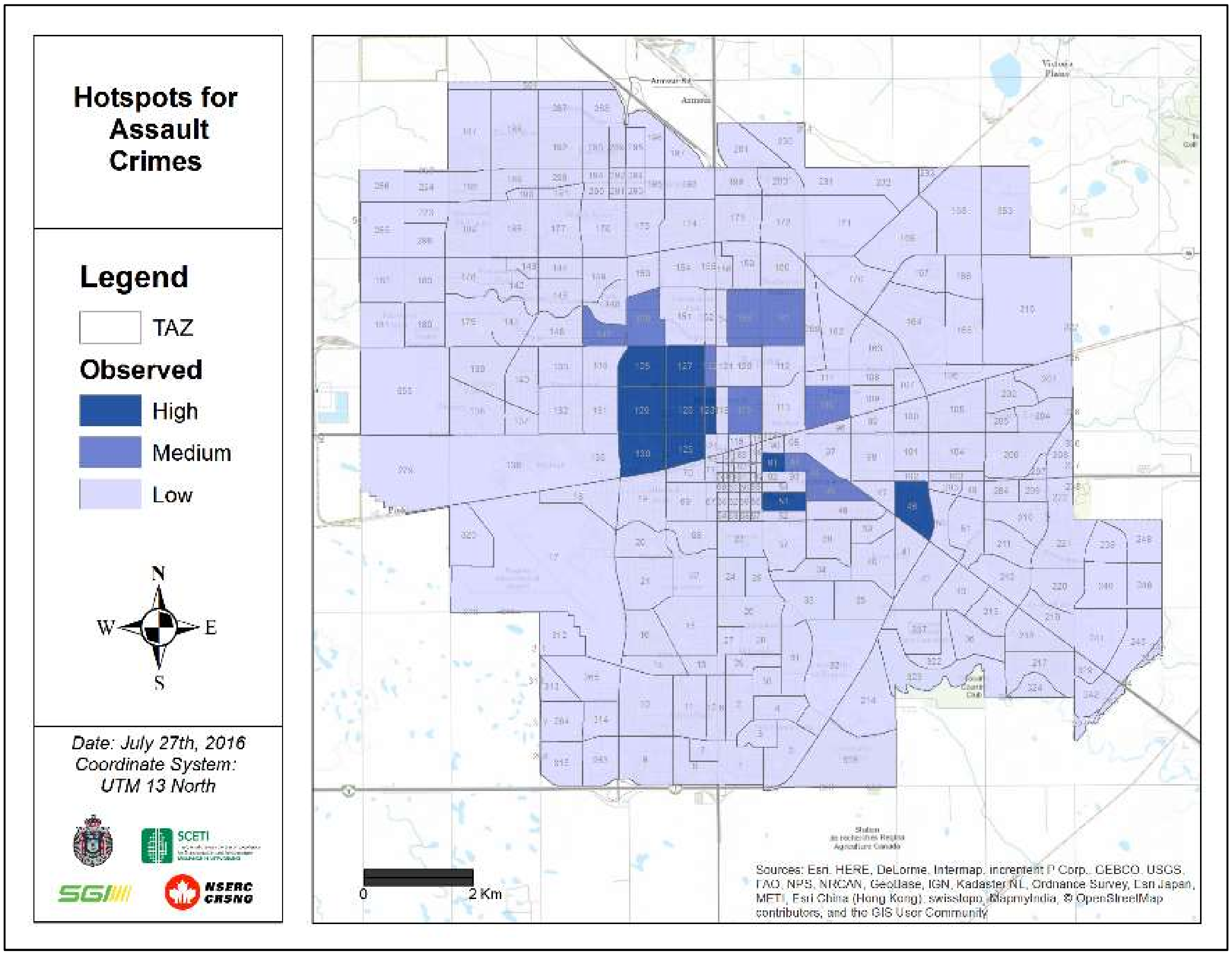
Variable	Observations	Total	Min.	Max.	Mean	Standard Deviation
Five-Violent Crimes	263	18231.8	0	9115.9	69.06	563.37

Observed Arson Crimes



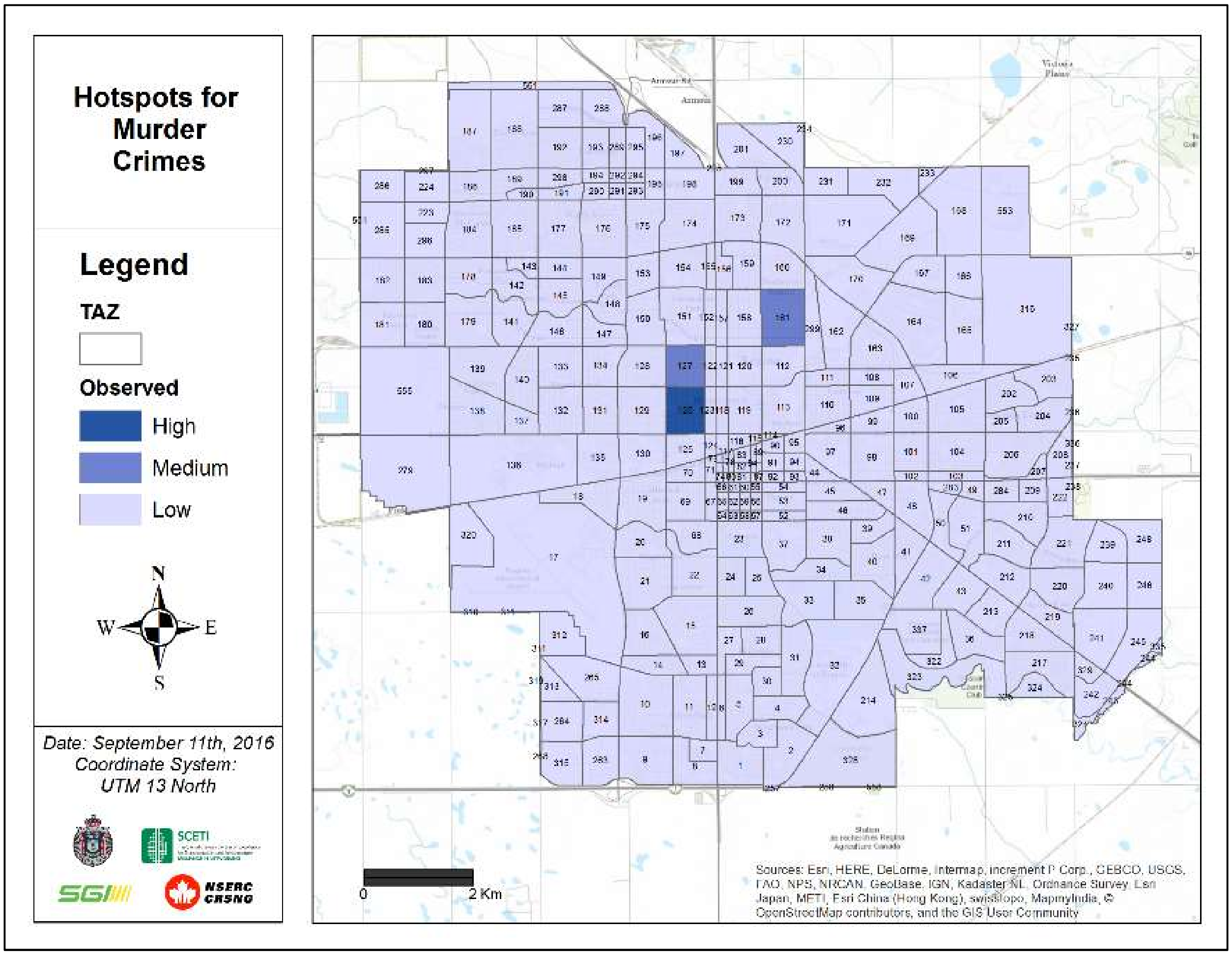
Variable	Observations	Total	Min.	Max.	Mean	Standard Deviation
Arson Crimes	263	268.7	0	60.3	1.02	4.06

Observed Assault Crimes



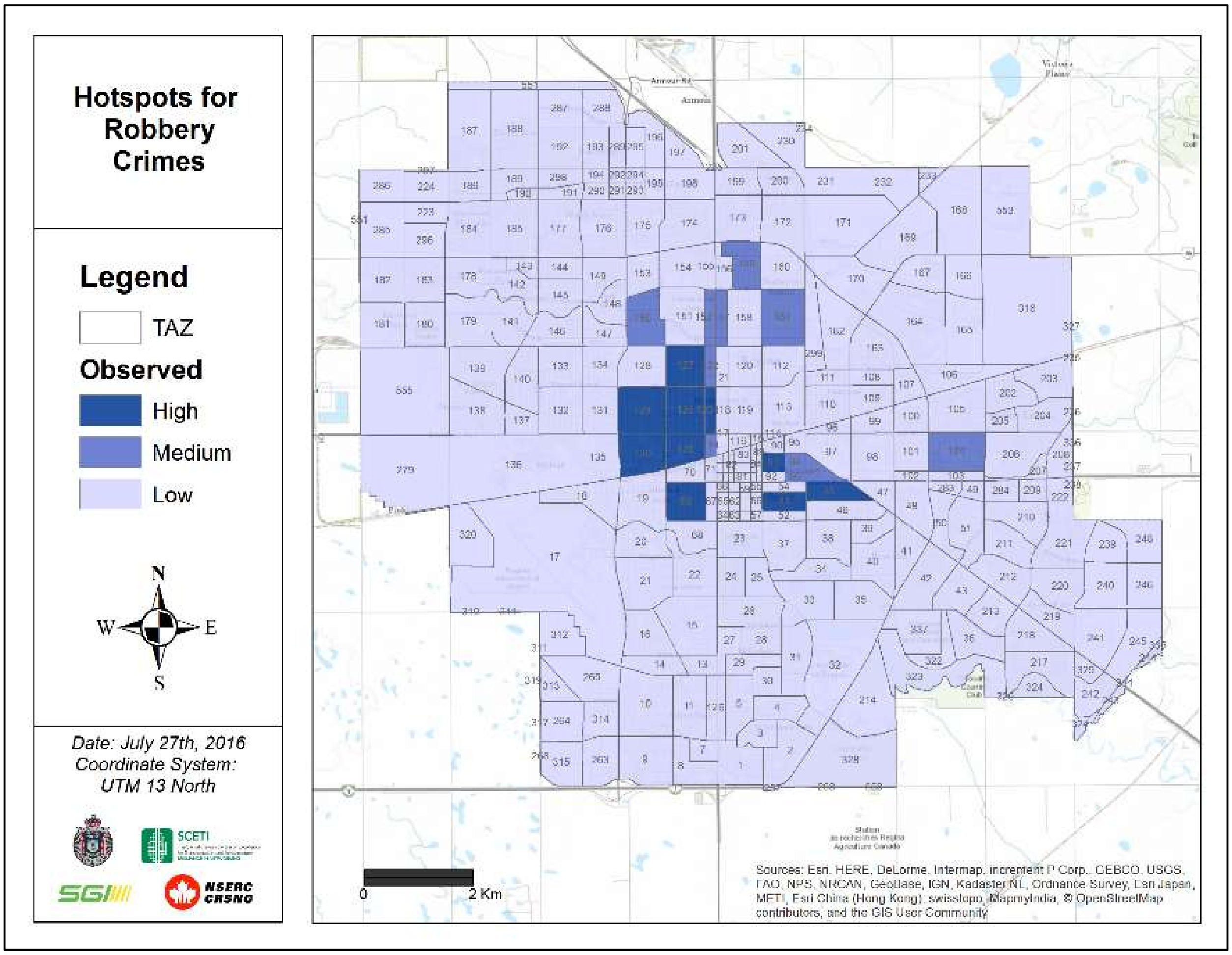
Variable	Observations	Total	Min.	Max.	Mean	Standard Deviation
Assault Crimes	263	6906.8	0	683.5	26.26	58.13

Observed Murder Crimes



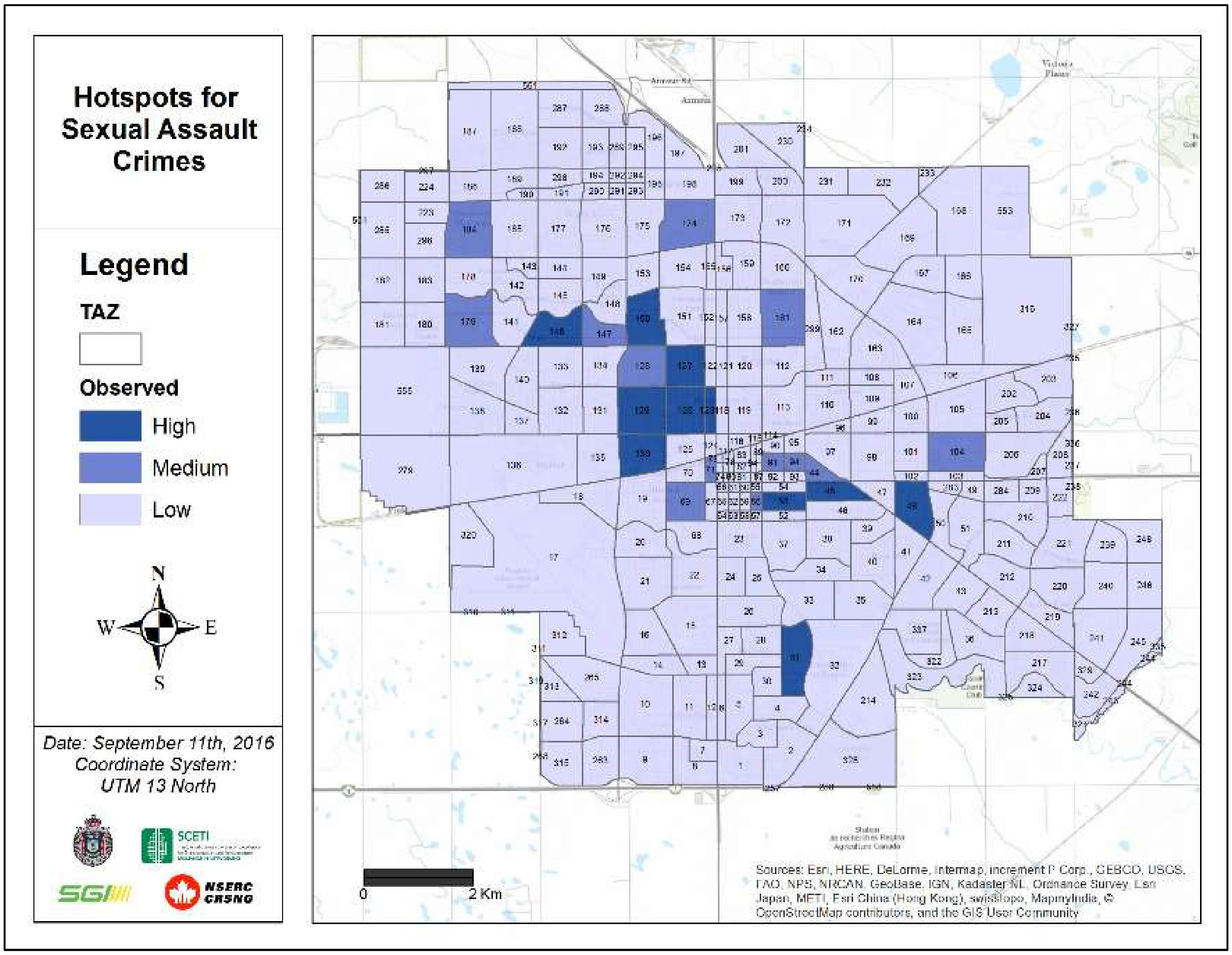
Variable	Observations	Total	Min.	Max.	Mean	Standard Deviation
Murder Crimes	263	24	0	7	0.09	0.52

Observed Robbery Crimes



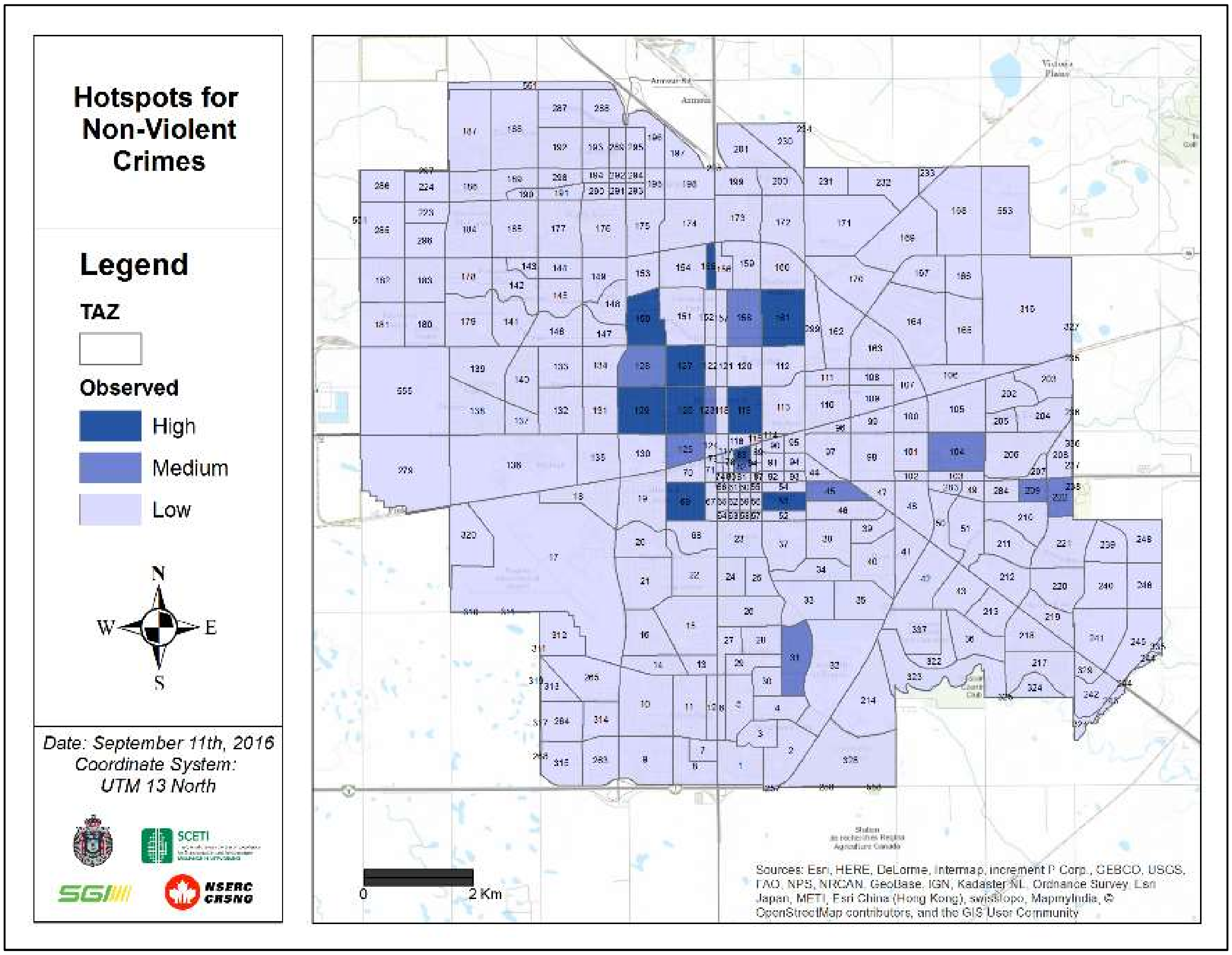
Variable	Observations	Total	Min.	Max.	Mean	Standard Deviation
Robbery Crimes	263	1300.6	0	141	4.95	12.18

Observed Sexual Assault Crimes



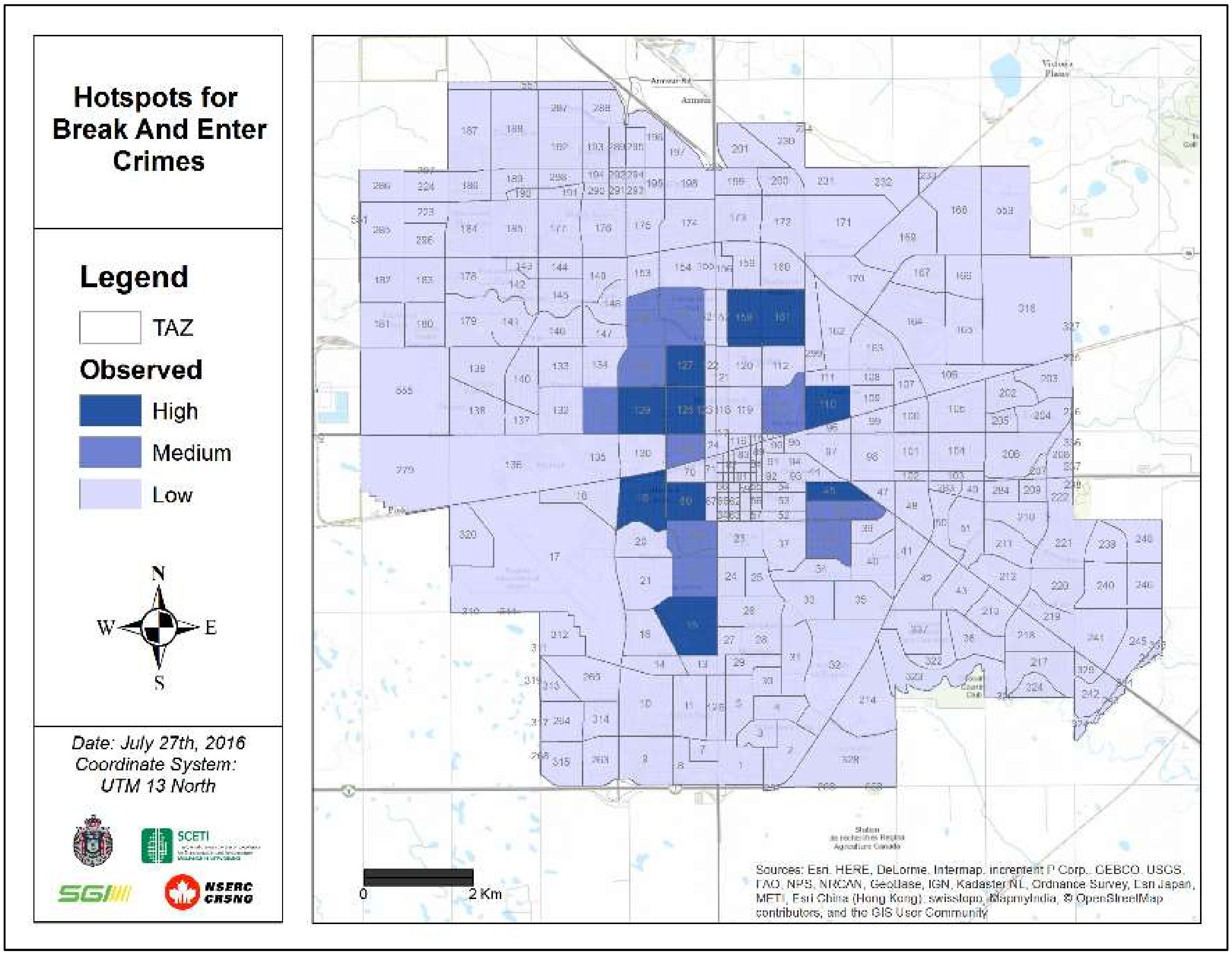
Variable	Observations	Total	Min.	Max.	Mean	Standard Deviation
Sexual Assault Crimes	263	609	0	66	2.32	6.20

Observed Non-Violent Crimes



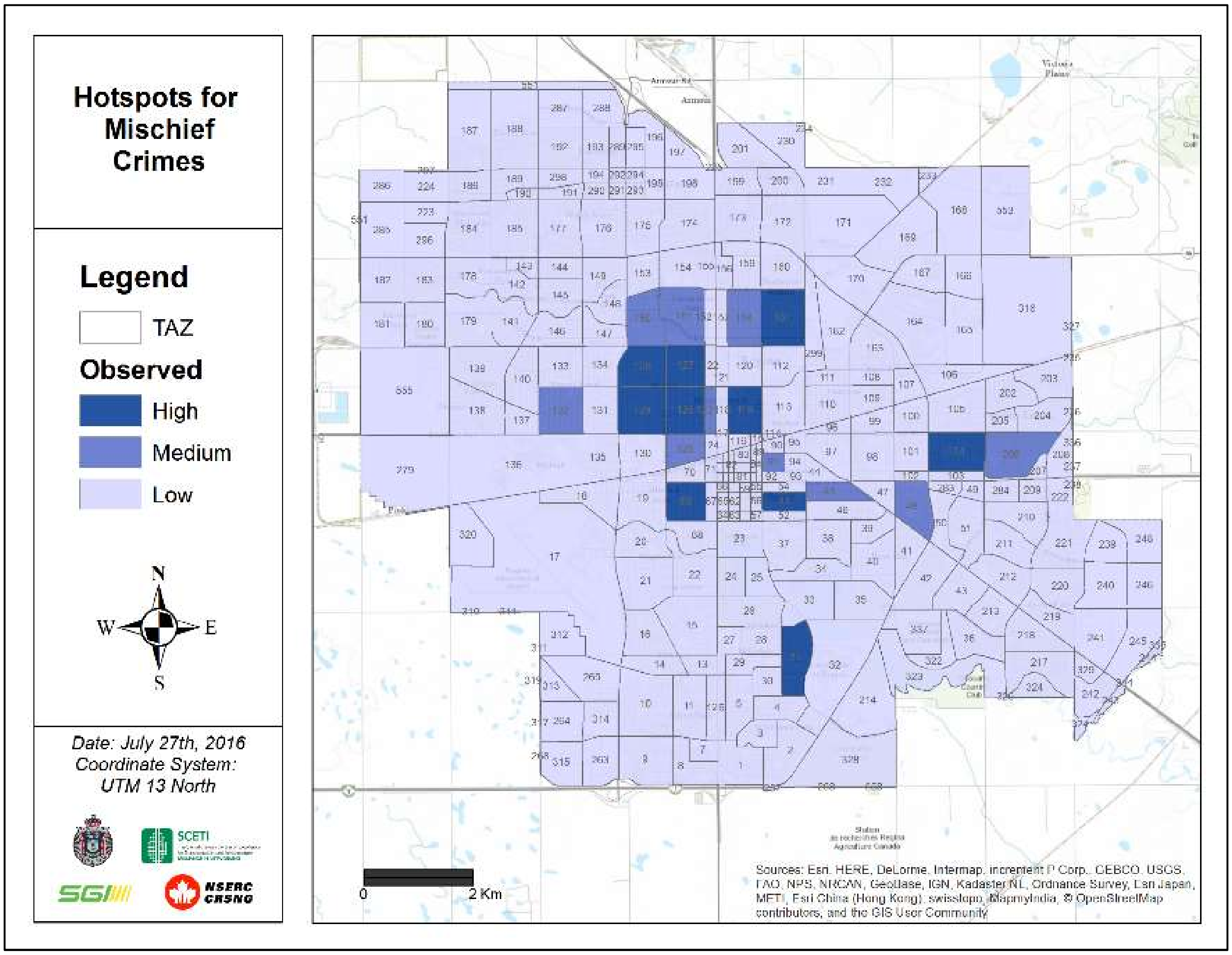
Variable	Observations	Total	Min.	Max.	Mean	Standard Deviation
Sexual Assault Crimes	263	41035.20	0	1479	141.99	177.61

Observed Break and Enter Crimes



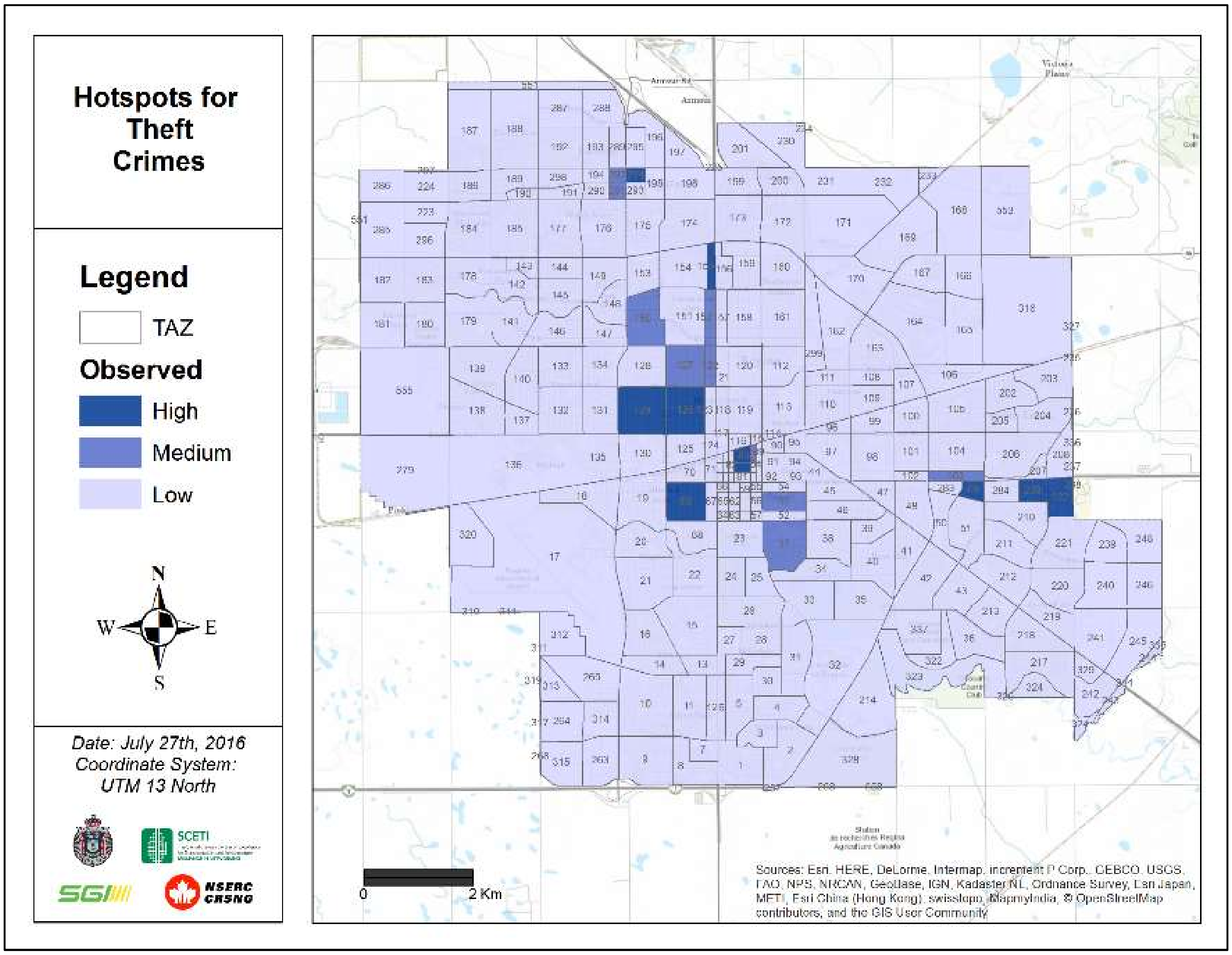
Variable	Observations	Total	Min.	Max.	Mean	Standard Deviation
Break and Enter Crimes	263	6057.6	0	290.5	23.03	32.27

Observed Mischief Crimes



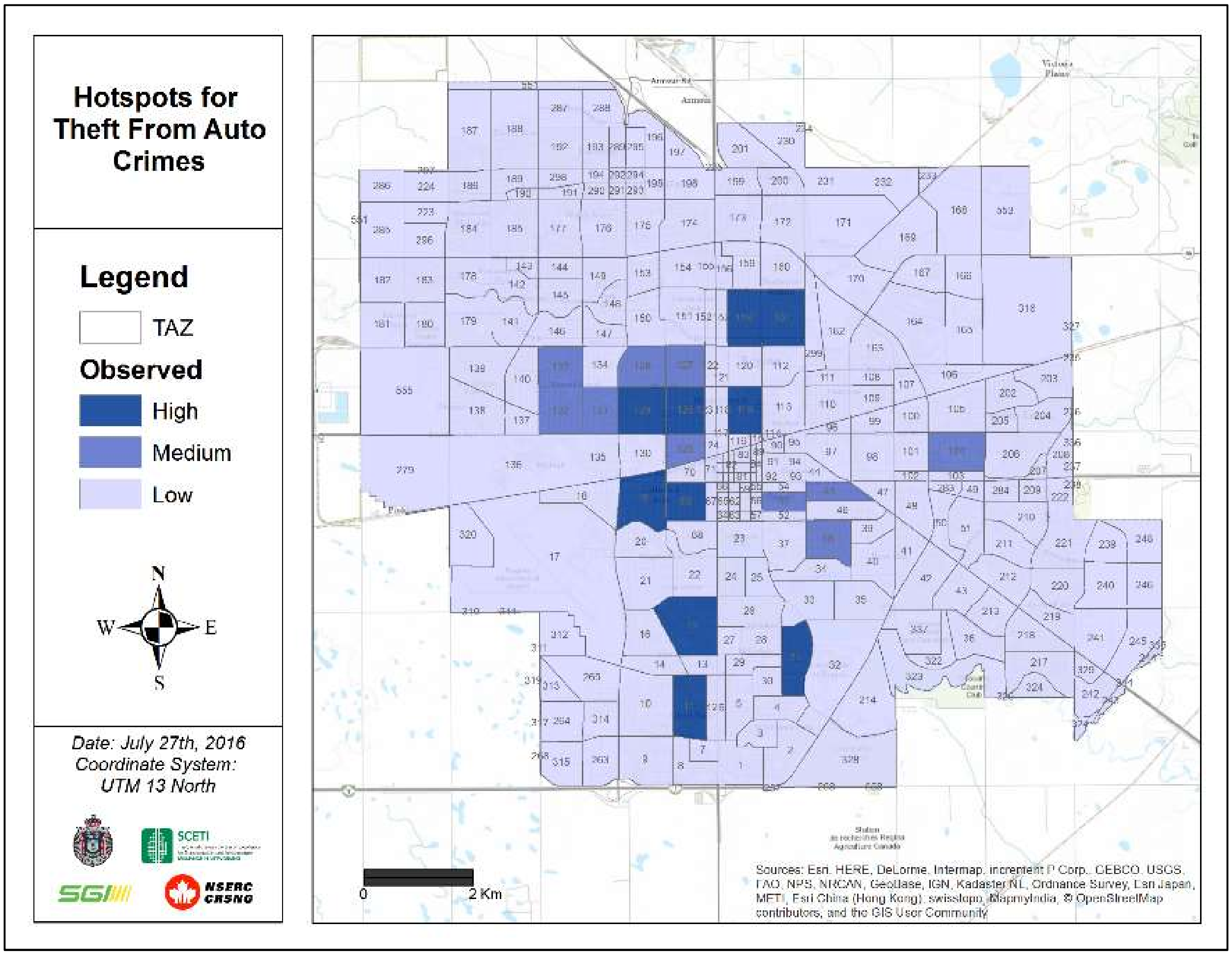
Variable	Observations	Total	Min.	Max.	Mean	Standard Deviation
Mischief Crimes	263	10905	0	530	41.46	54.91

Observed Theft Crimes



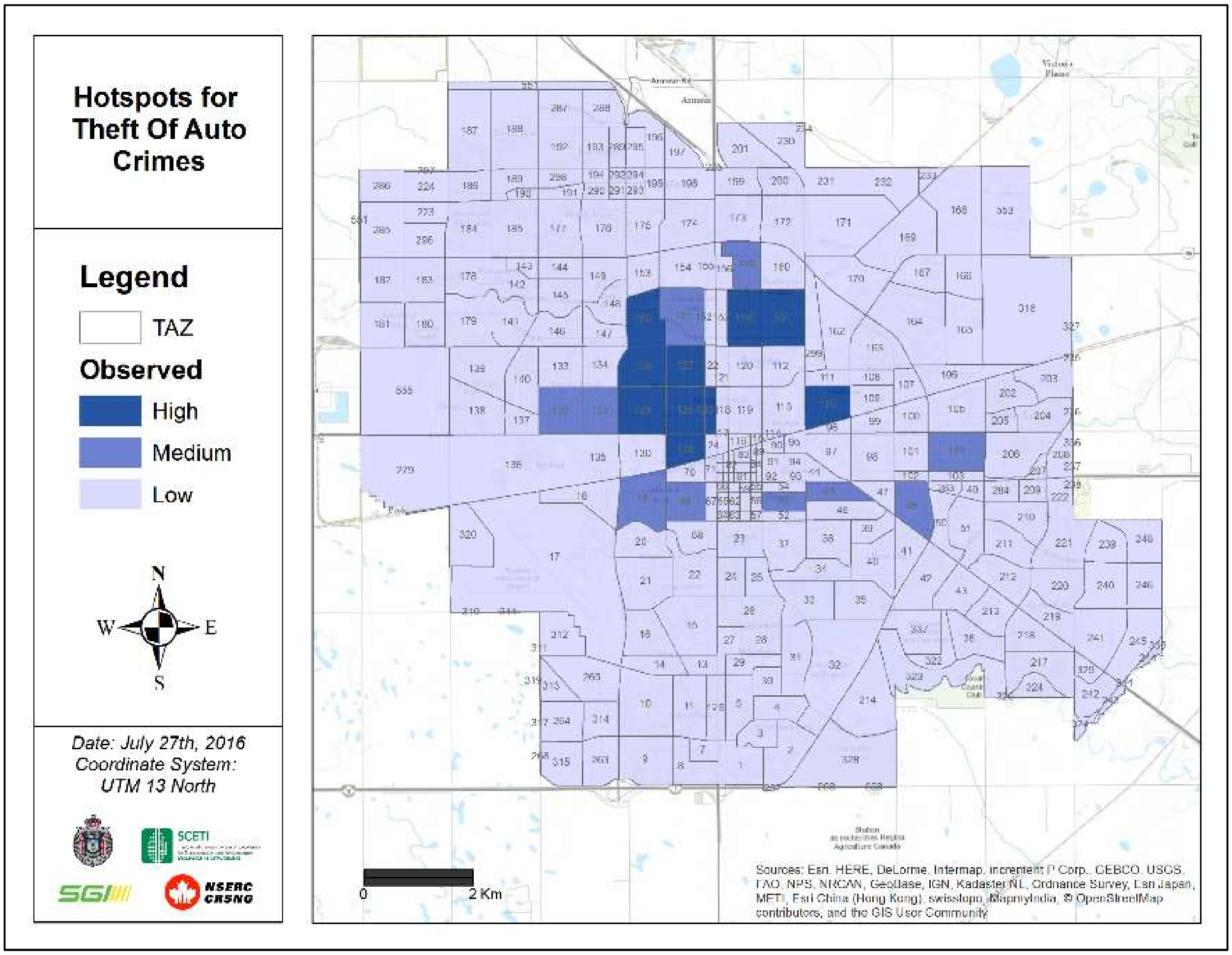
Variable	Observations	Total	Min.	Max.	Mean	Standard Deviation
Theft Crimes	263	11421.2	0	593.2	43.43	77.78

Observed Theft from Auto Crimes



Variable	Observations	Total	Min.	Max.	Mean	Standard Deviation
Theft from Auto Crimes	263	8037.6	0	174.8	30.56	30.67

Observed Theft of Auto Crimes

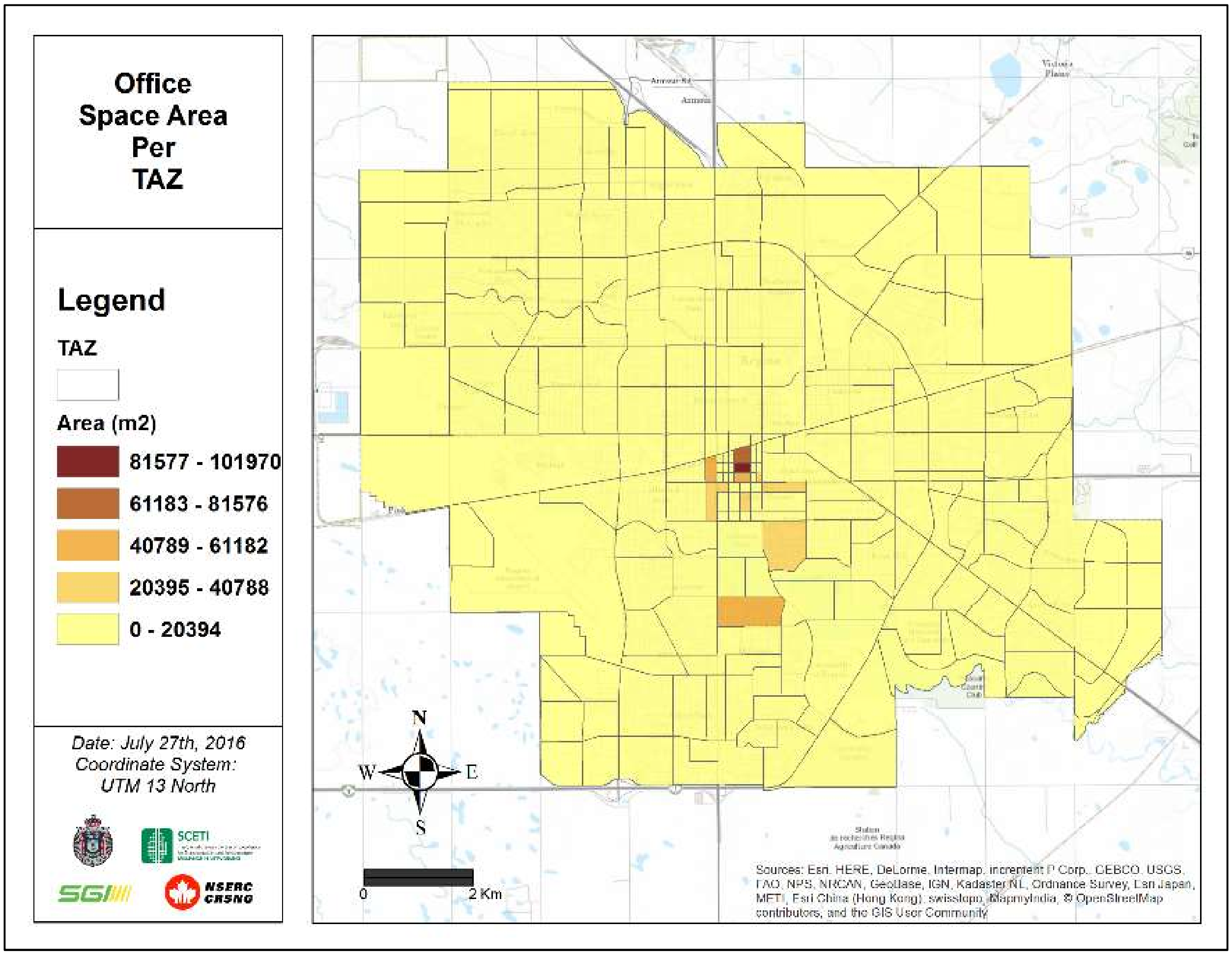


Variable	Observations	Total	Min.	Max.	Mean	Standard Deviation
Theft of Auto Crimes	263	4648.5	0	241	17.57	26.62

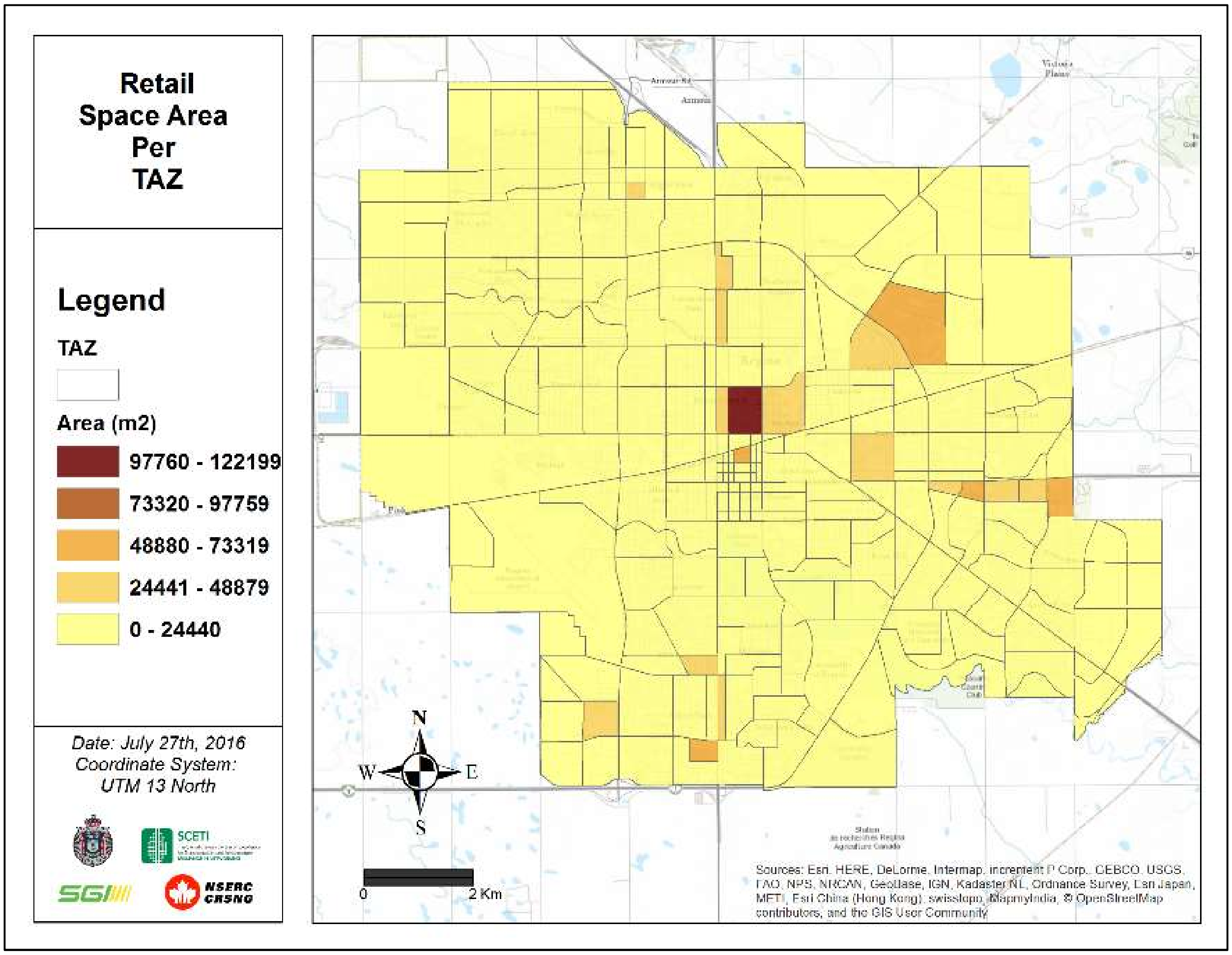
C3. Socio-Economic and Land Use Data

Variable	Observations	Total	Min.	Maximum	Mean	Standard Deviation
Office Space (per m^2)	263	902366	0	101970	3431.05	10266.25
Retail Space (per m^2)	263	1628686	0	122199	6192.72	13029.34
Industry Space (per m^2)	263	1613288	0	274205	6134.17	24540.63
Hospital Space (per m^2)	263	117231	0	66704	445.745	4732.10
Number of Land use per TAZ	263	1086	1	8	4.13	1.68
Commercial Area (m^2)	263	7403575	0	312623	28150.48	46800.05
Institutional Area (m^2)	263	4221201.88	0	620487.35	16050.20	46778.39
Open Space Recreational Area (m^2)	263	24674256.21	0	2501580.25	93818.46	248948.94
Railway Area (m^2)	263	2777417	0	691835	10560.52	52694.82
High Density Residential Area	263	4212293.98	0	196674.39	16016.33	32603.35
Low Density Residential Area (m^2)	263	45363260	0	786855	172483.88	210194.97
Medium Density Residential Area (m^2)	263	3664149	0	383555	13932.13	43650.73
Urban Holding Residential Area (m^2)	263	21963243	0	3150675.91	83510.43	322005.57
TAZ Area (m^2)	263	137132121.34	0	6044326.67	521414.91	570457.13
Population Density (sq. km)	263	527664.94	0	10552.611	2006.34	1665.67
Residential Area (m^2)	263	75202944.78729	0	3140675.91	285942.76	364716.78

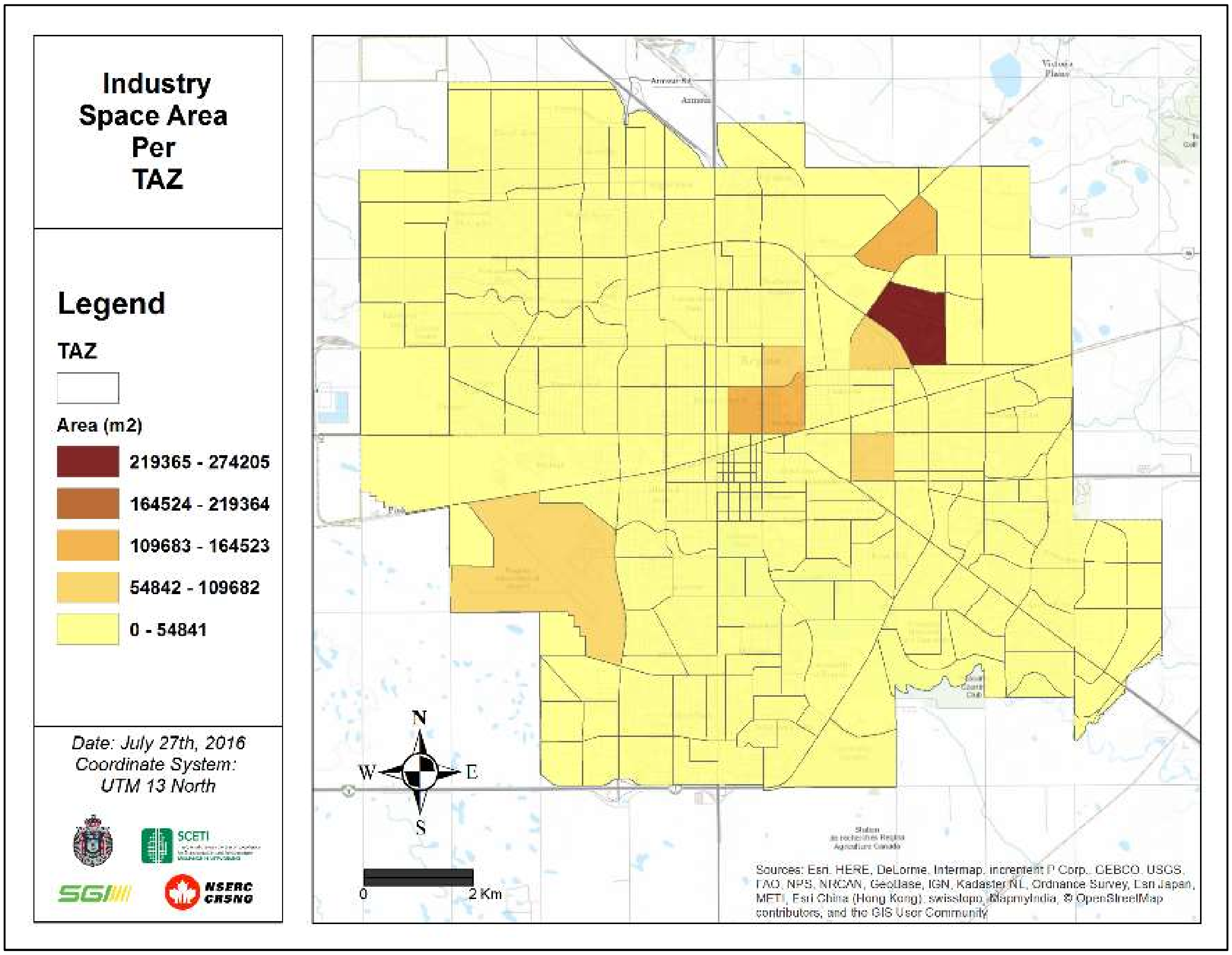
Office Space Area Land Use (m^2)



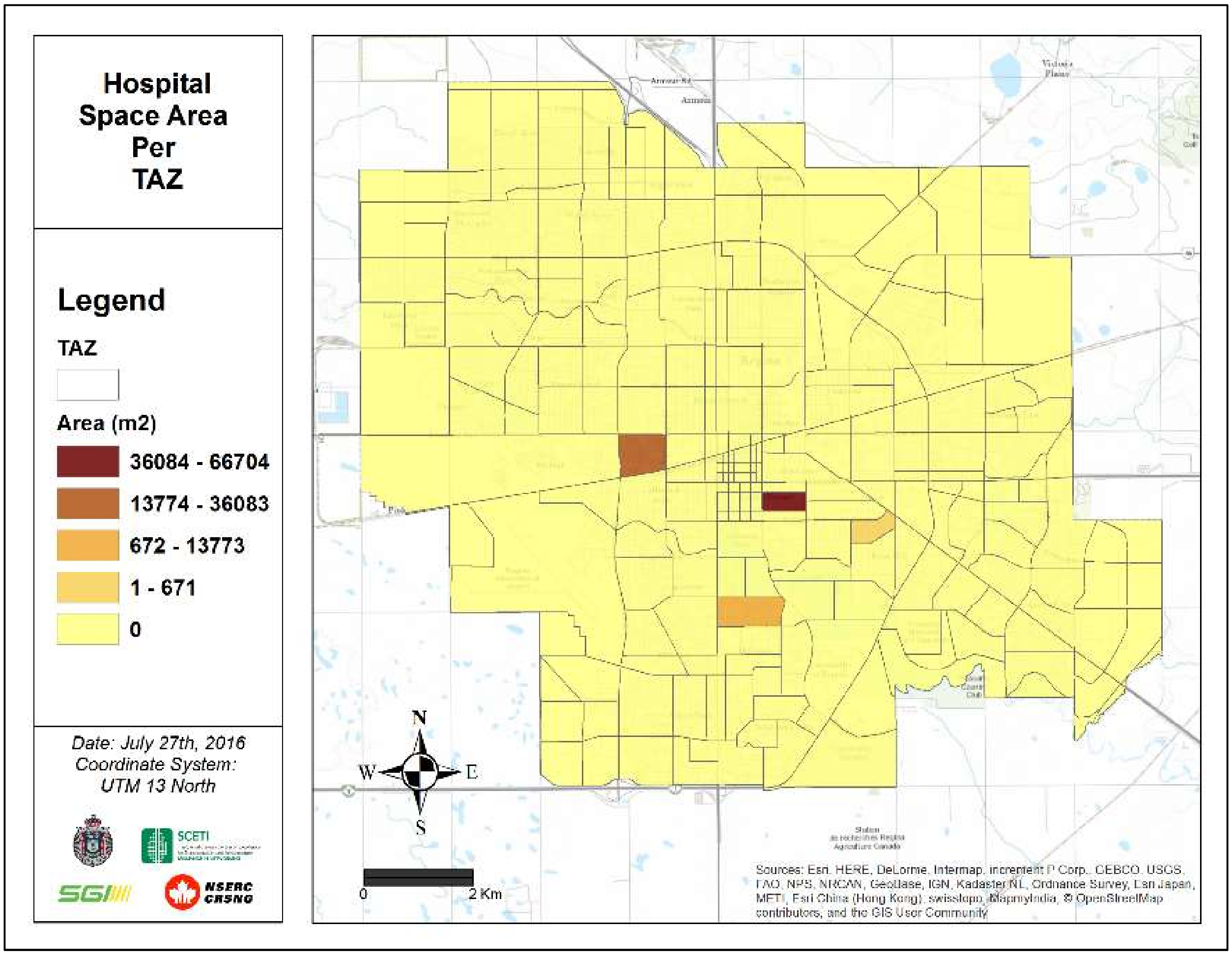
Retail Space Area Land Use (m^2)



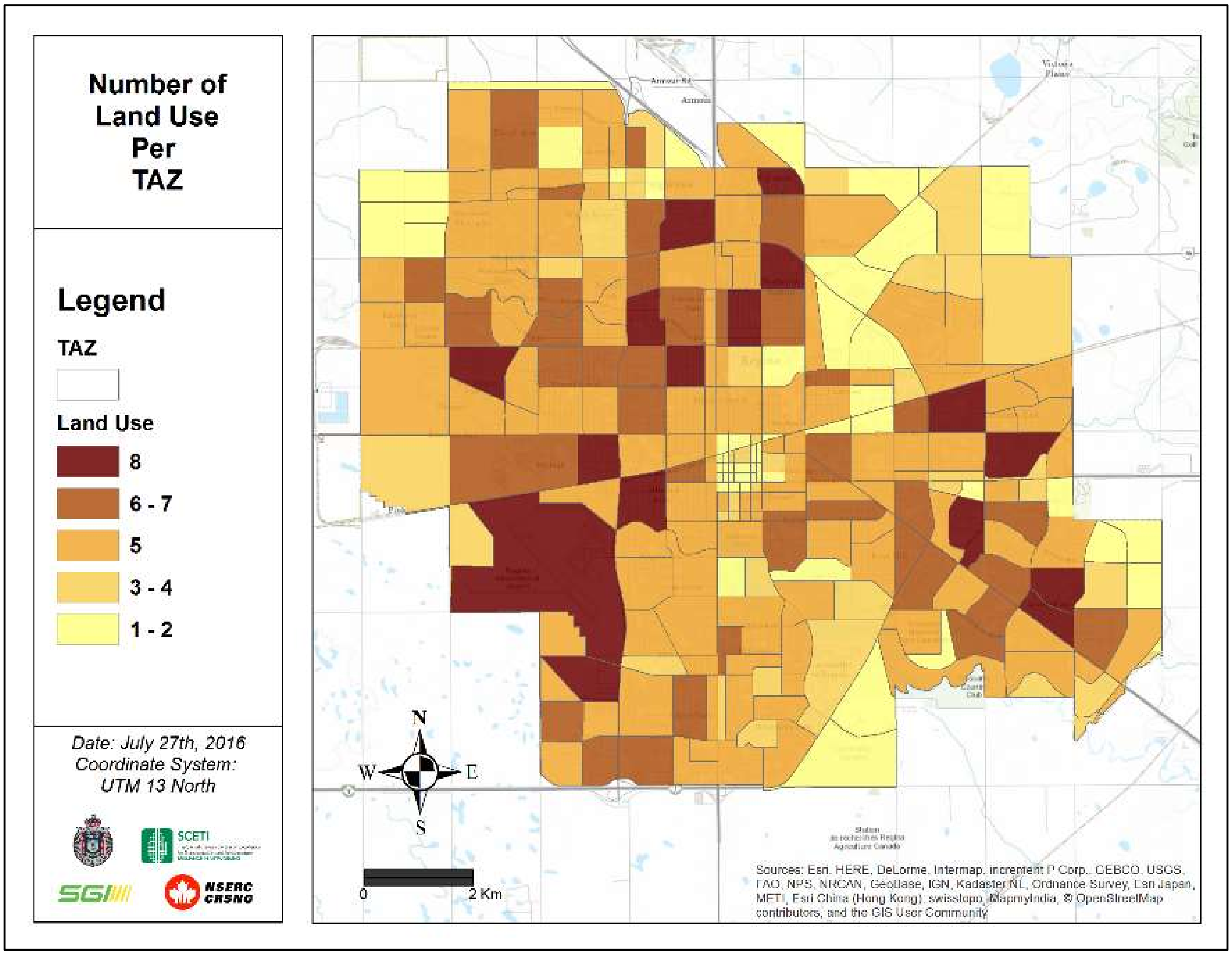
Industry Space Area Land Use (m^2)



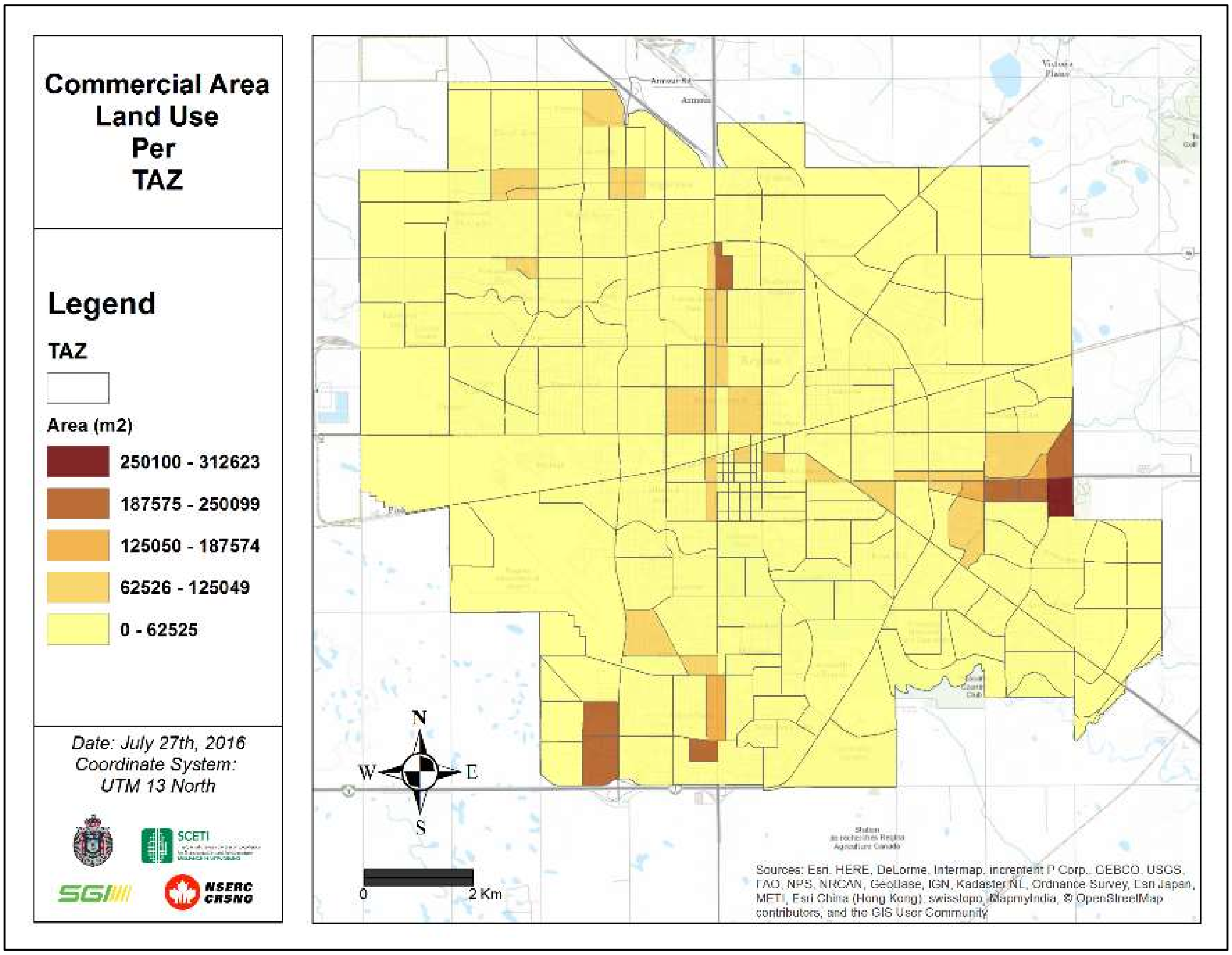
Hospital Space Area Land Use (m^2)



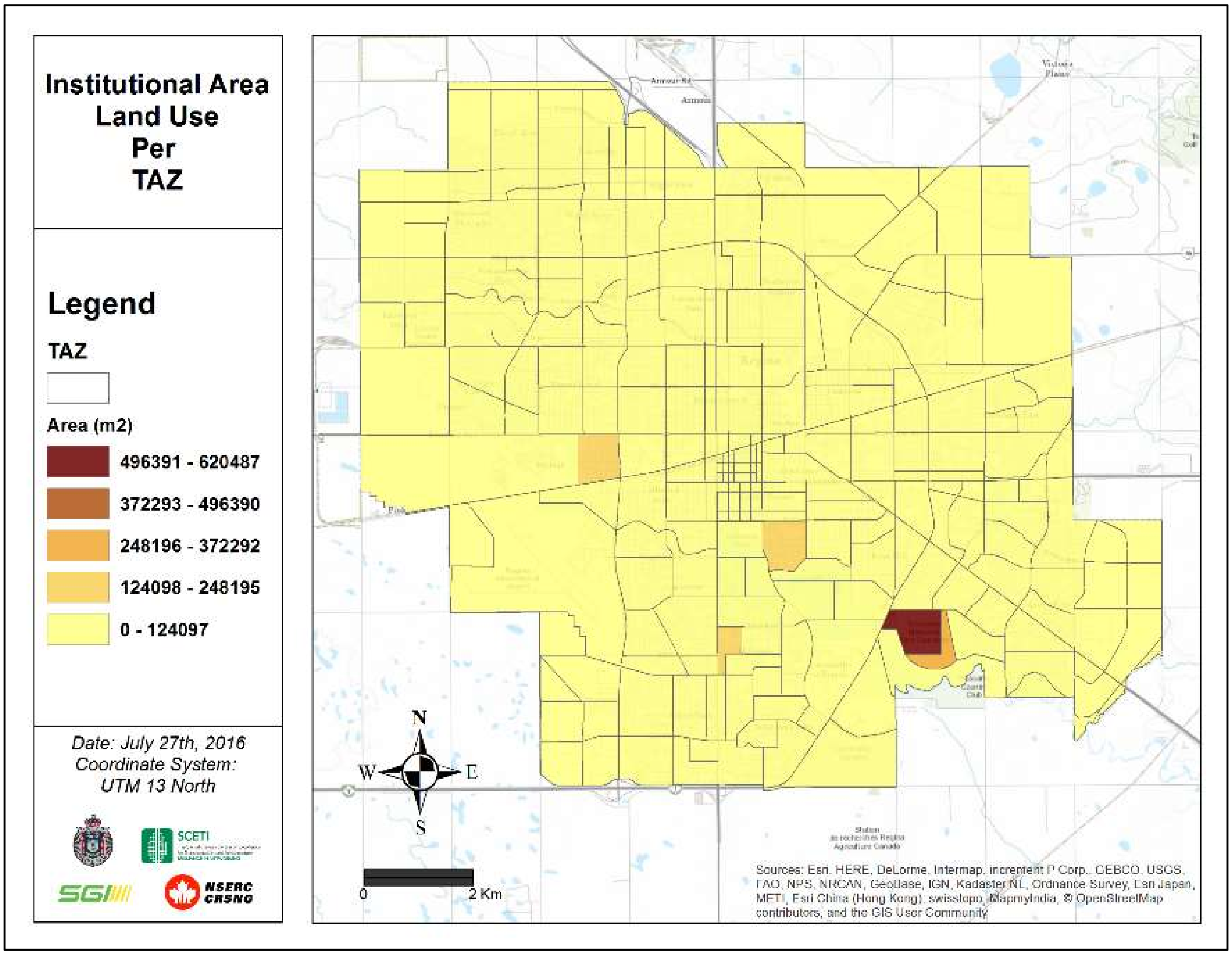
Number of Land Use



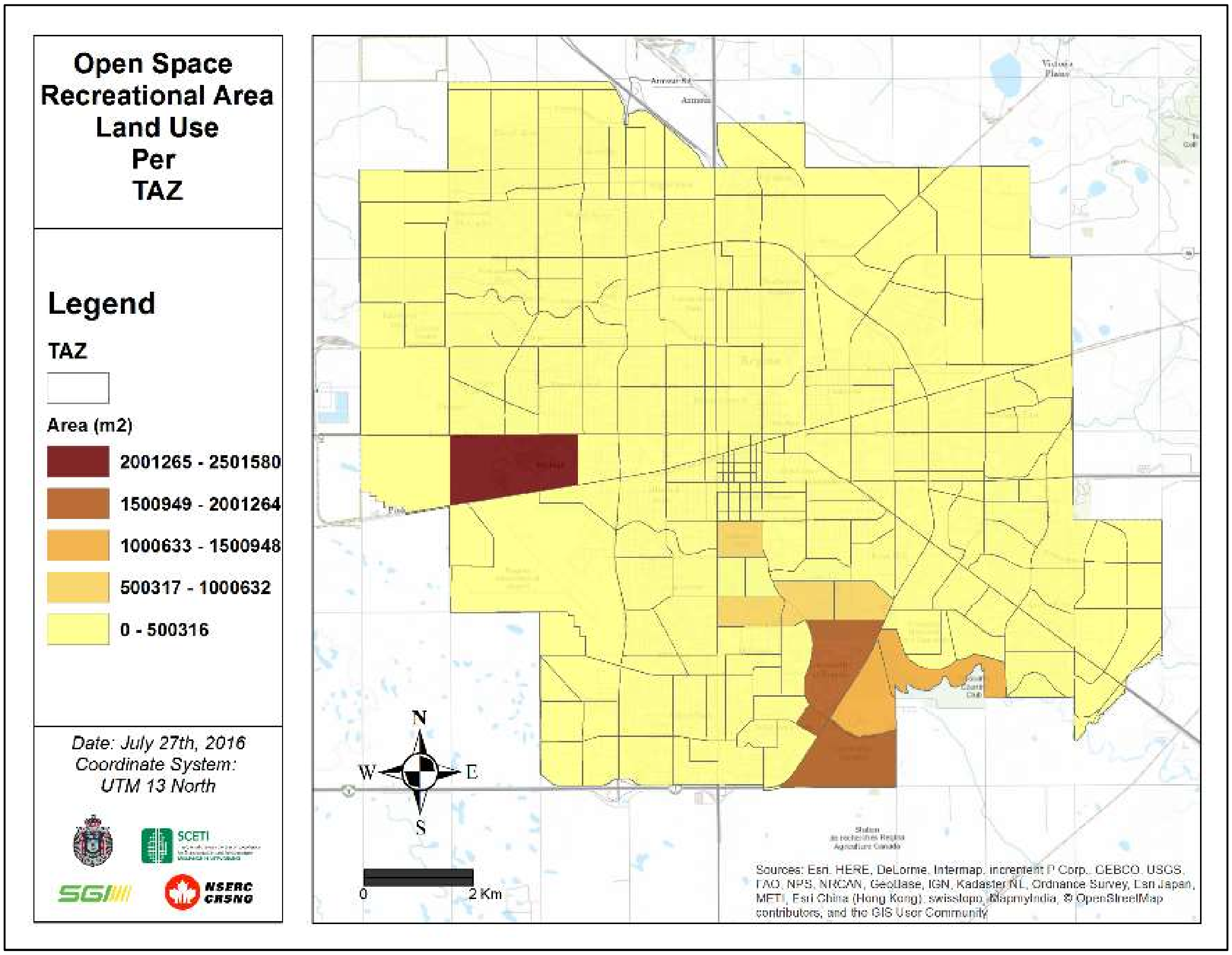
Commercial Area Land Use (m^2)



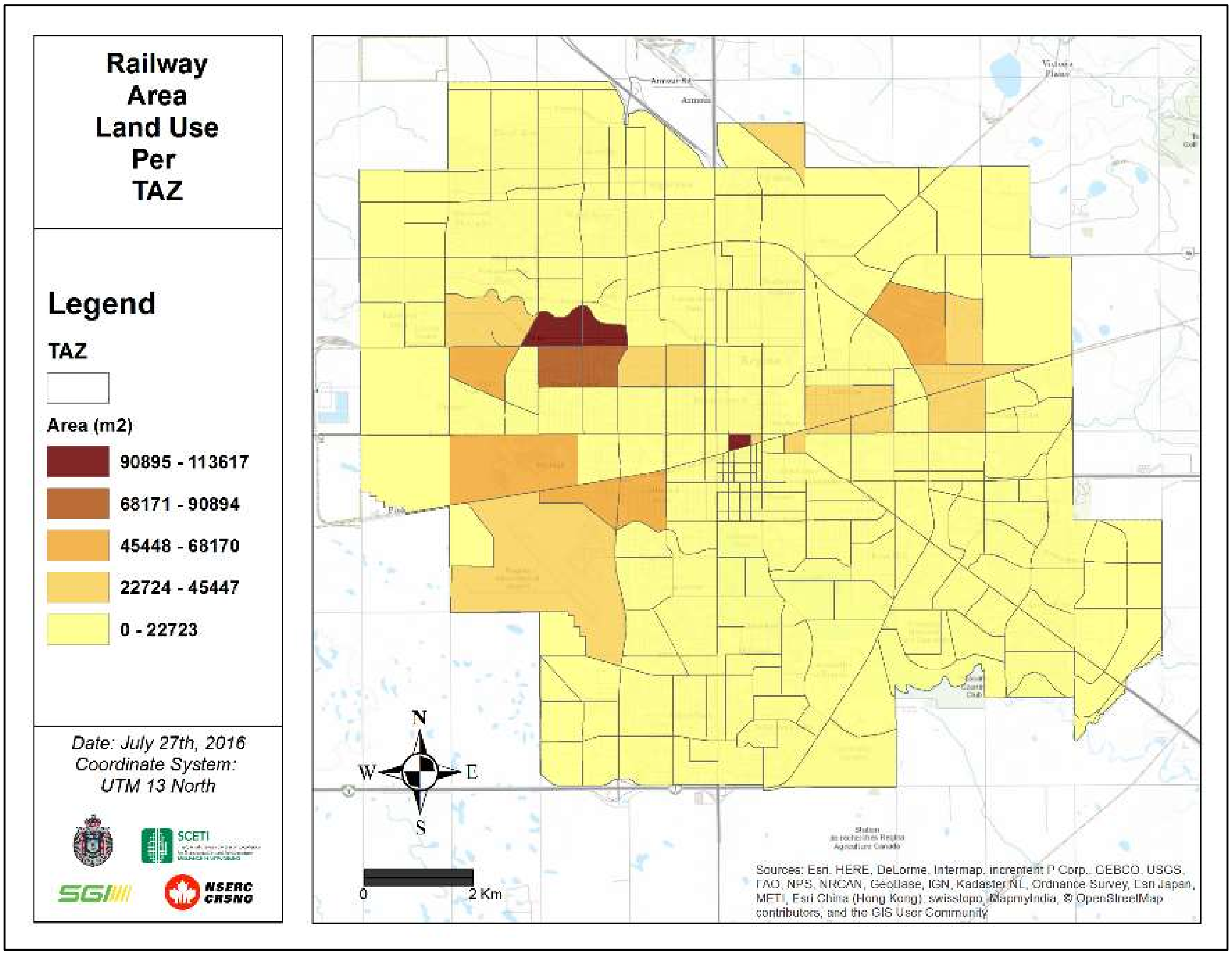
Institutional Area Land Use (m^2)



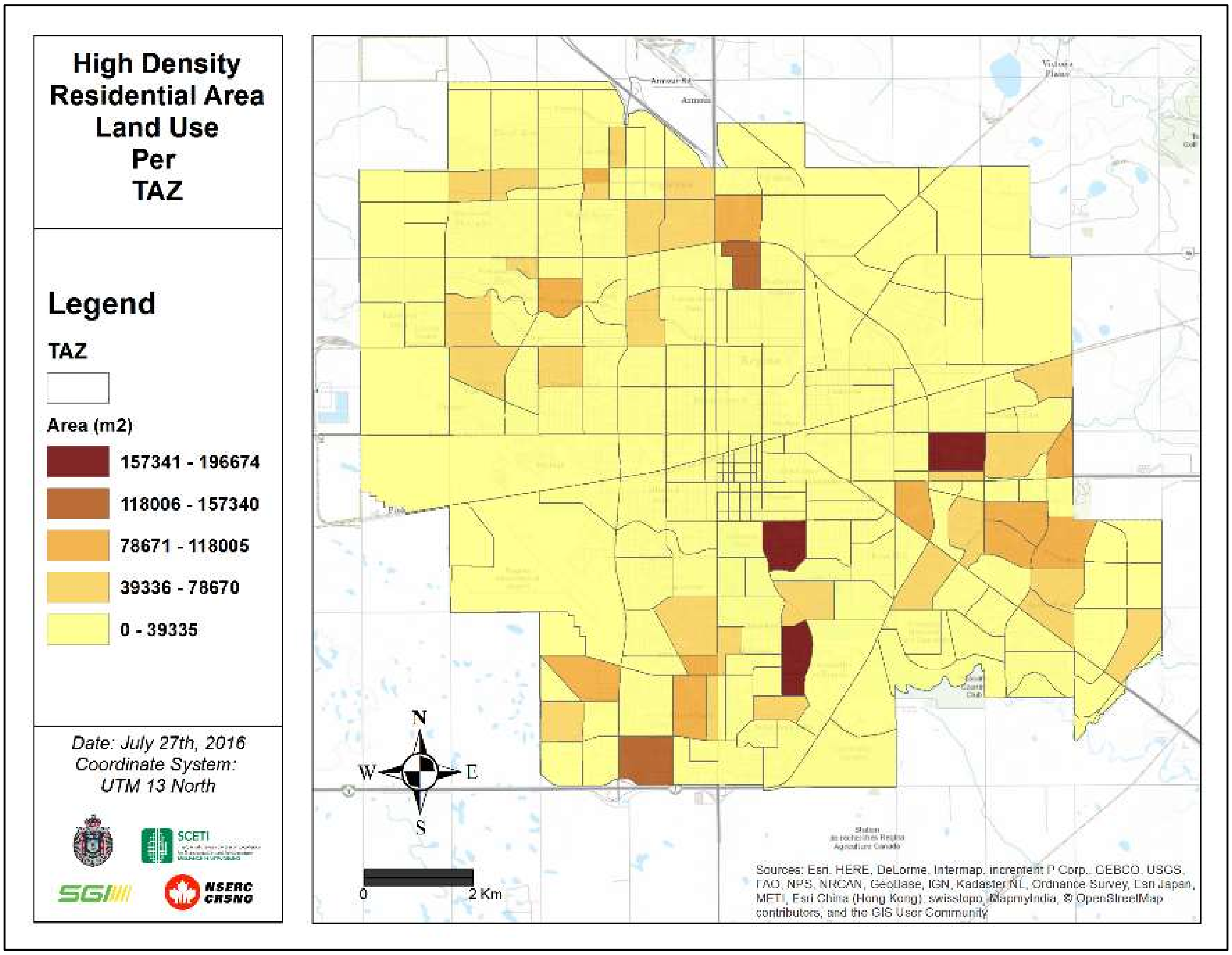
Open Space Recreational Area Land Use (m^2)



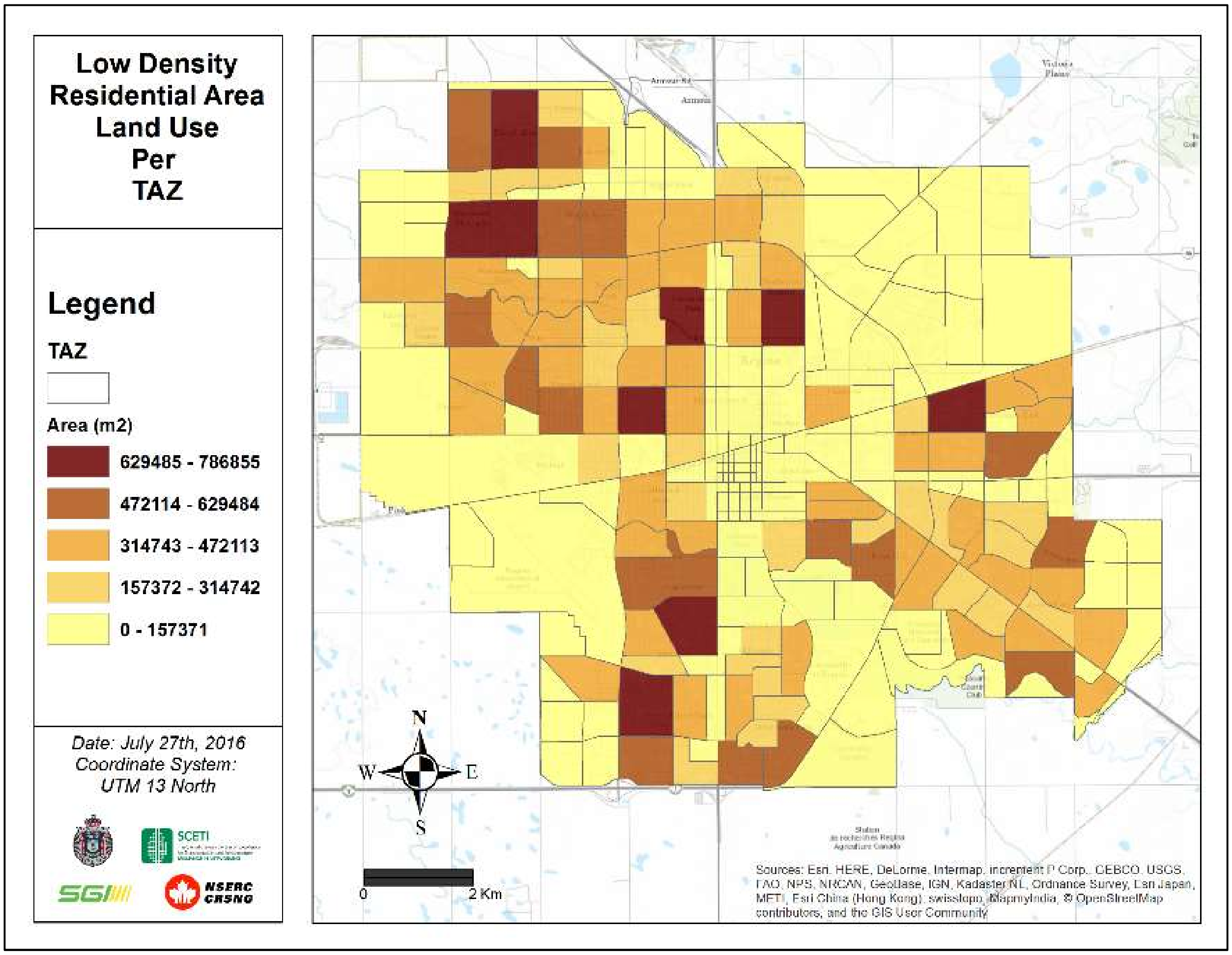
Railway Area Land Use (m^2)



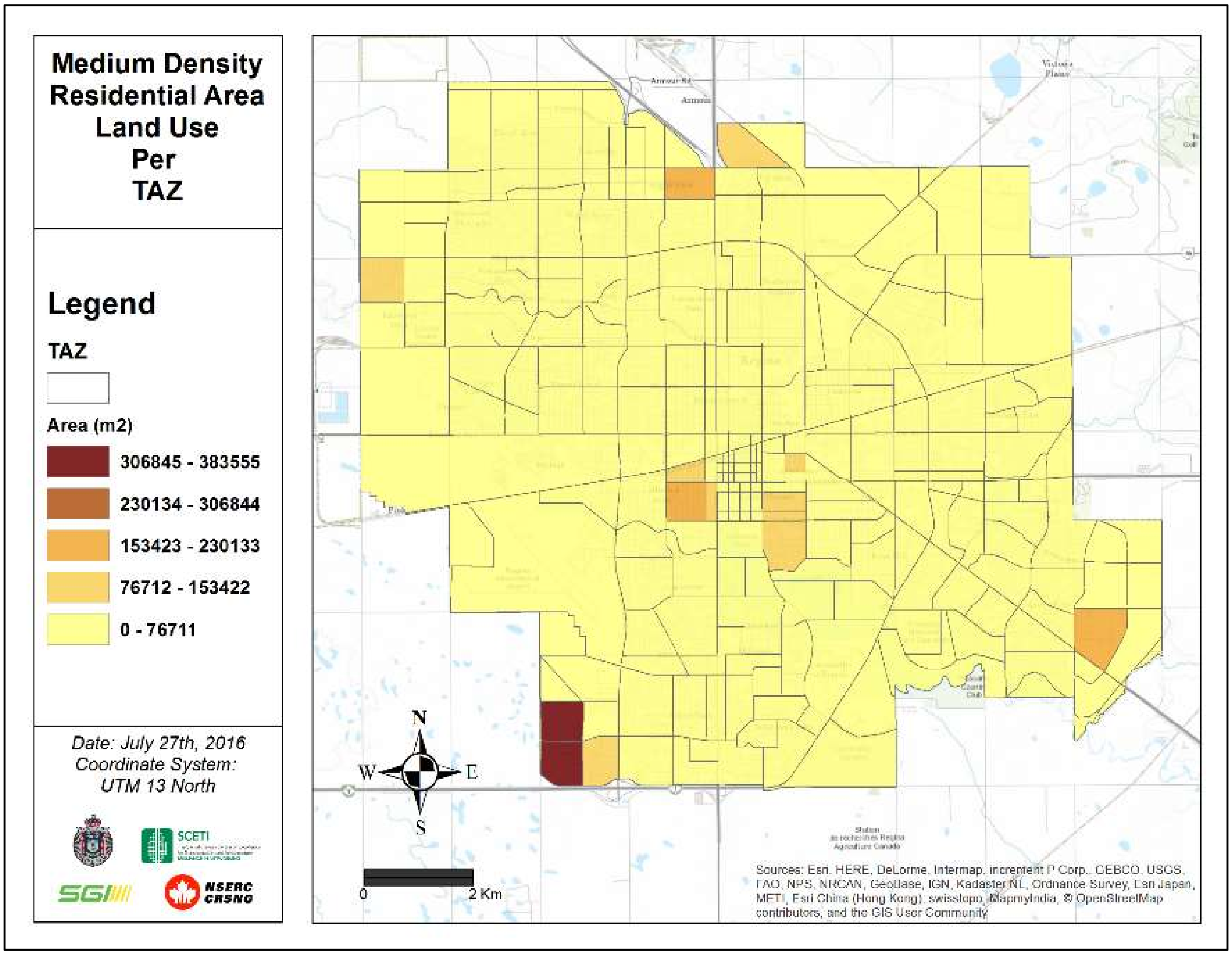
High Density Residential Area Land Use (m^2)



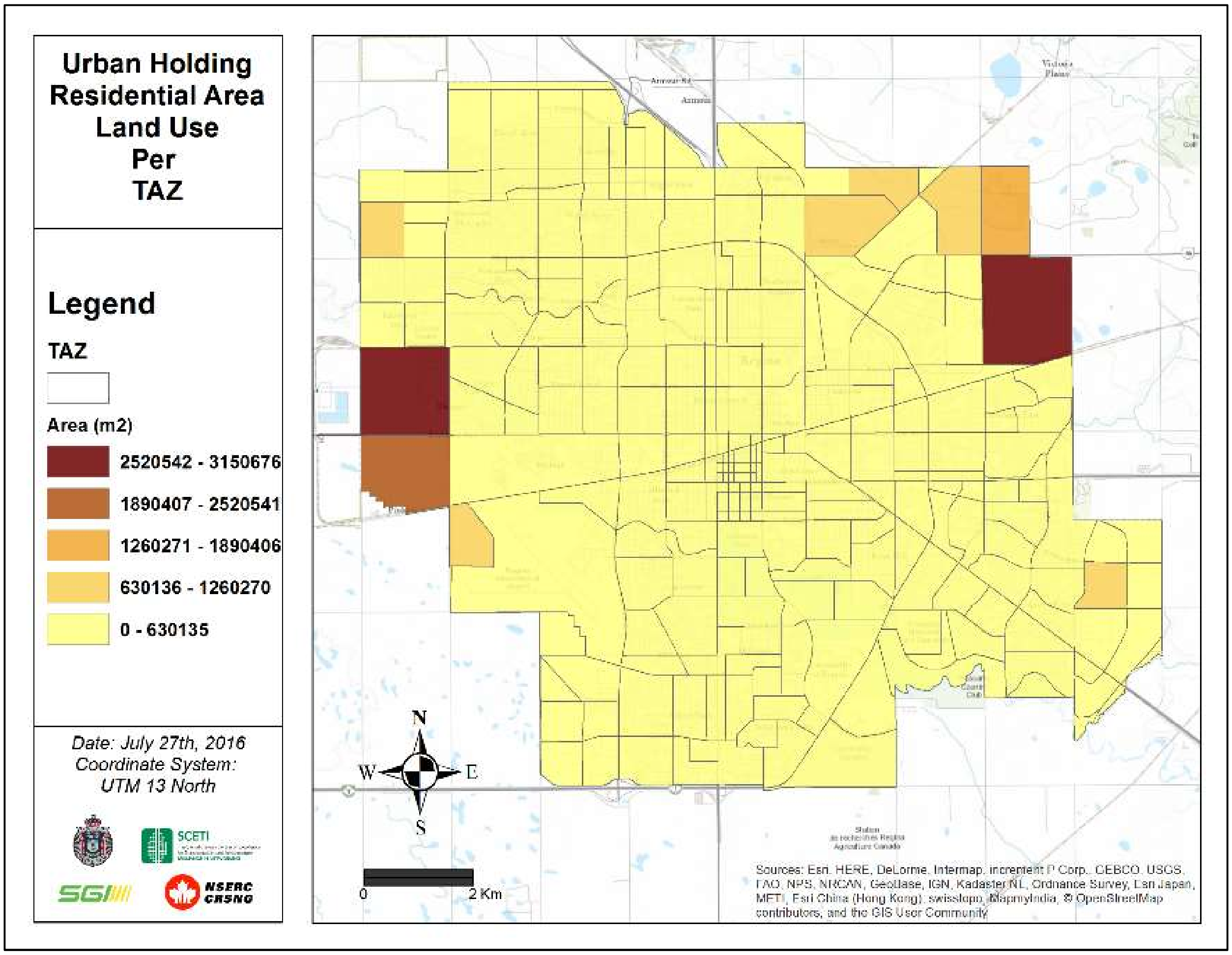
Low Density Residential Area Land Use (m^2)



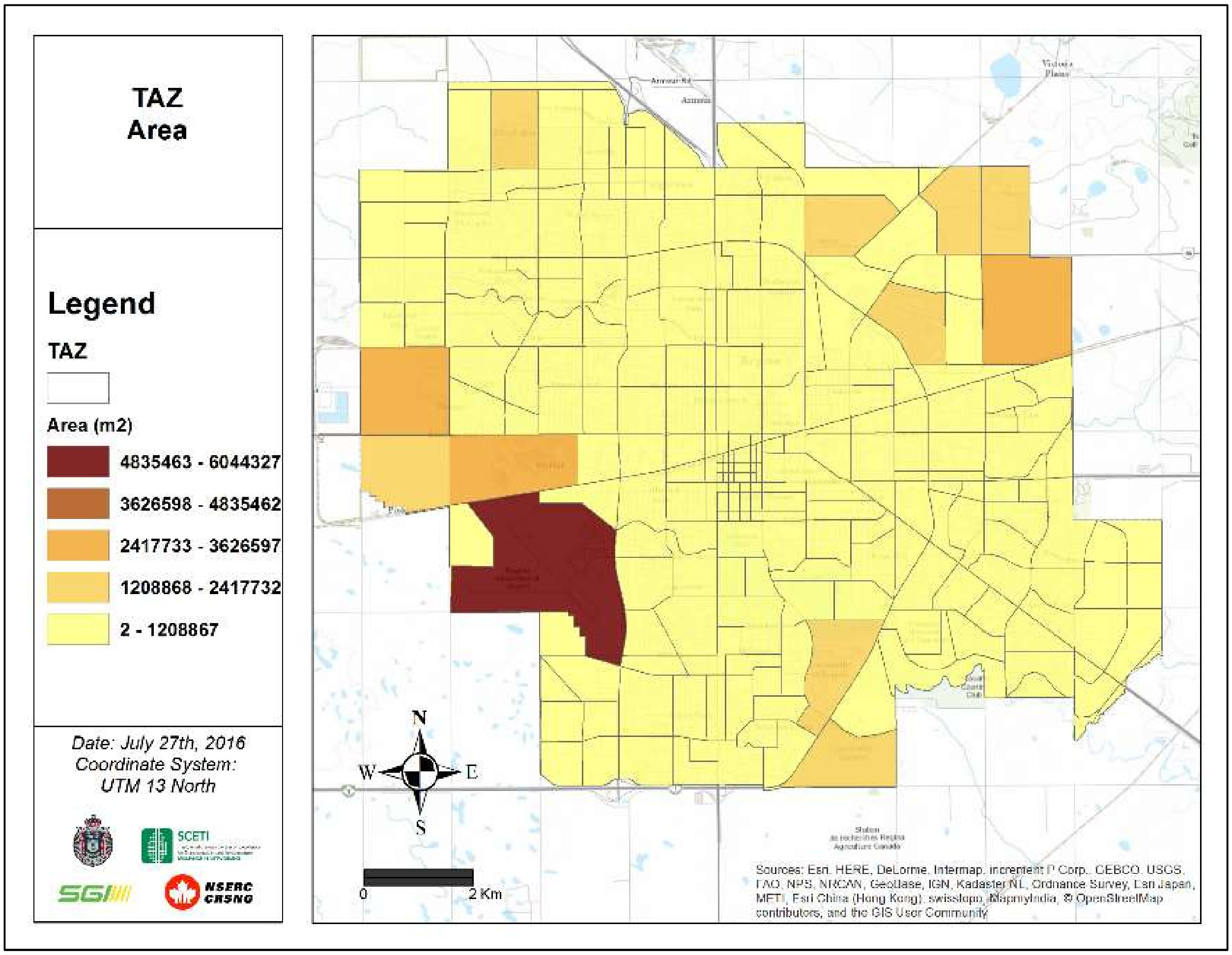
Medium Density Residential Area Land Use (m^2)



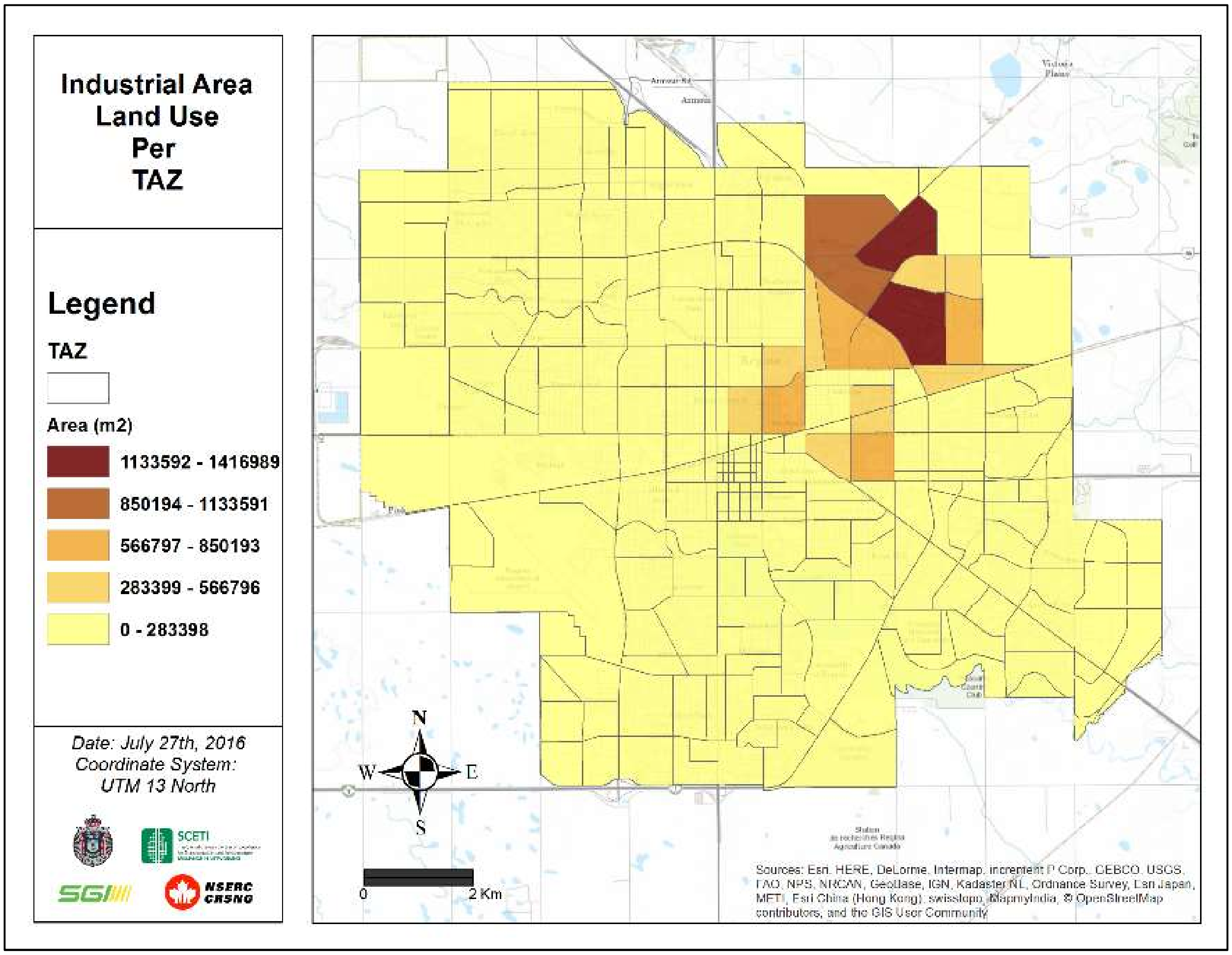
Urban Holding Residential Area Land Use (m^2)



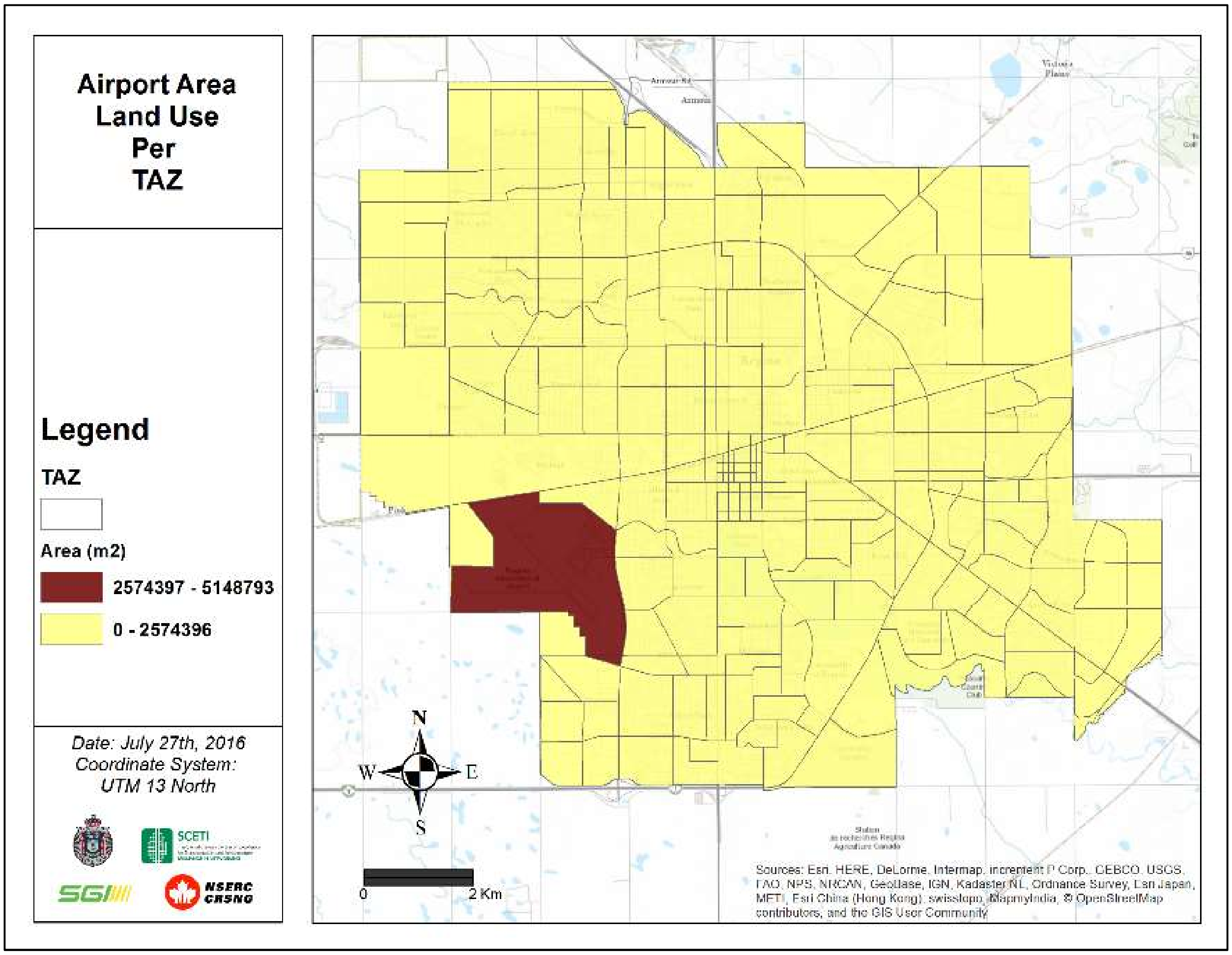
Traffic Analysis Zone (TAZ) Area (m^2)



Industrial Area Land Use (m^2)



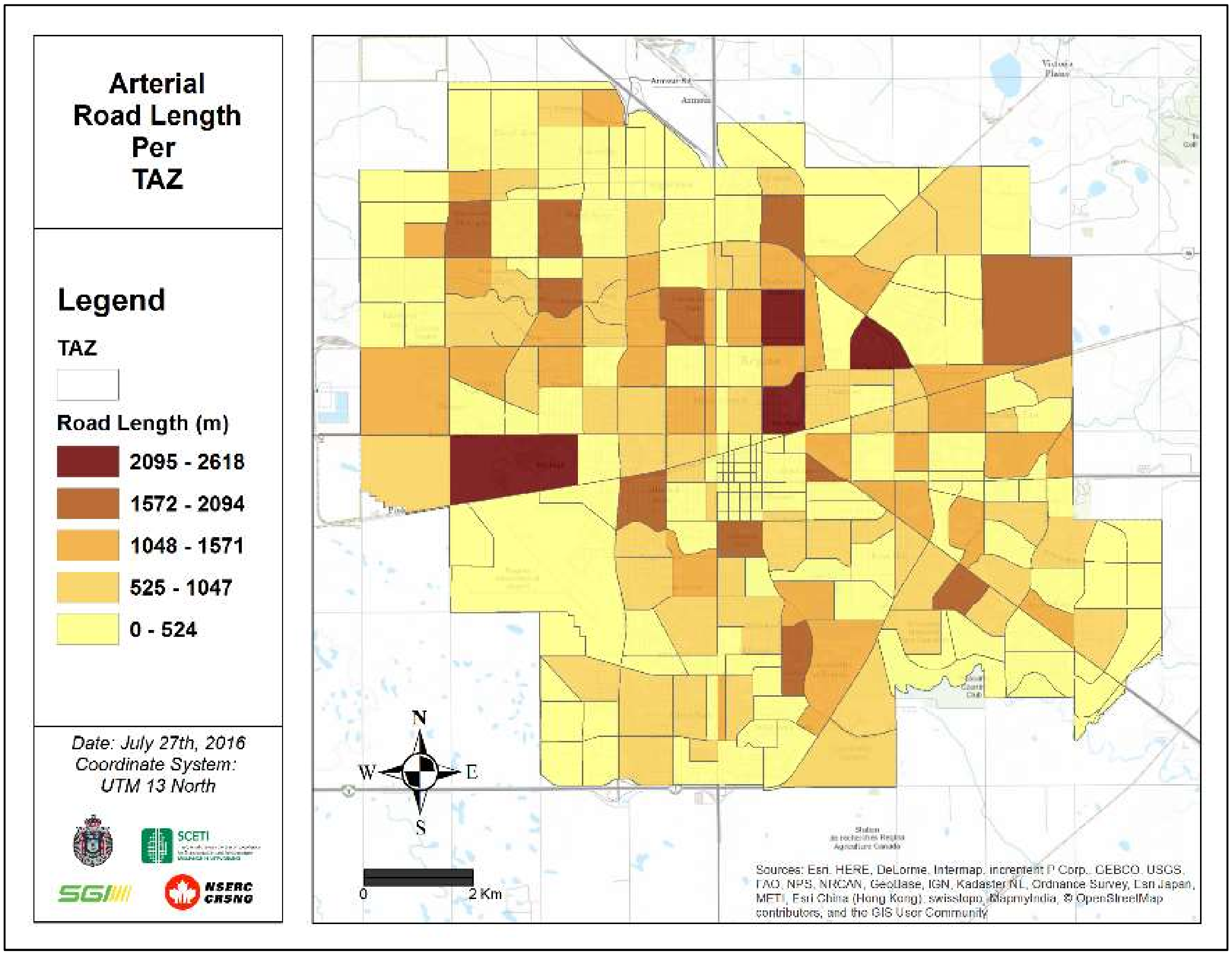
Airport Space Area Land Use (m^2)



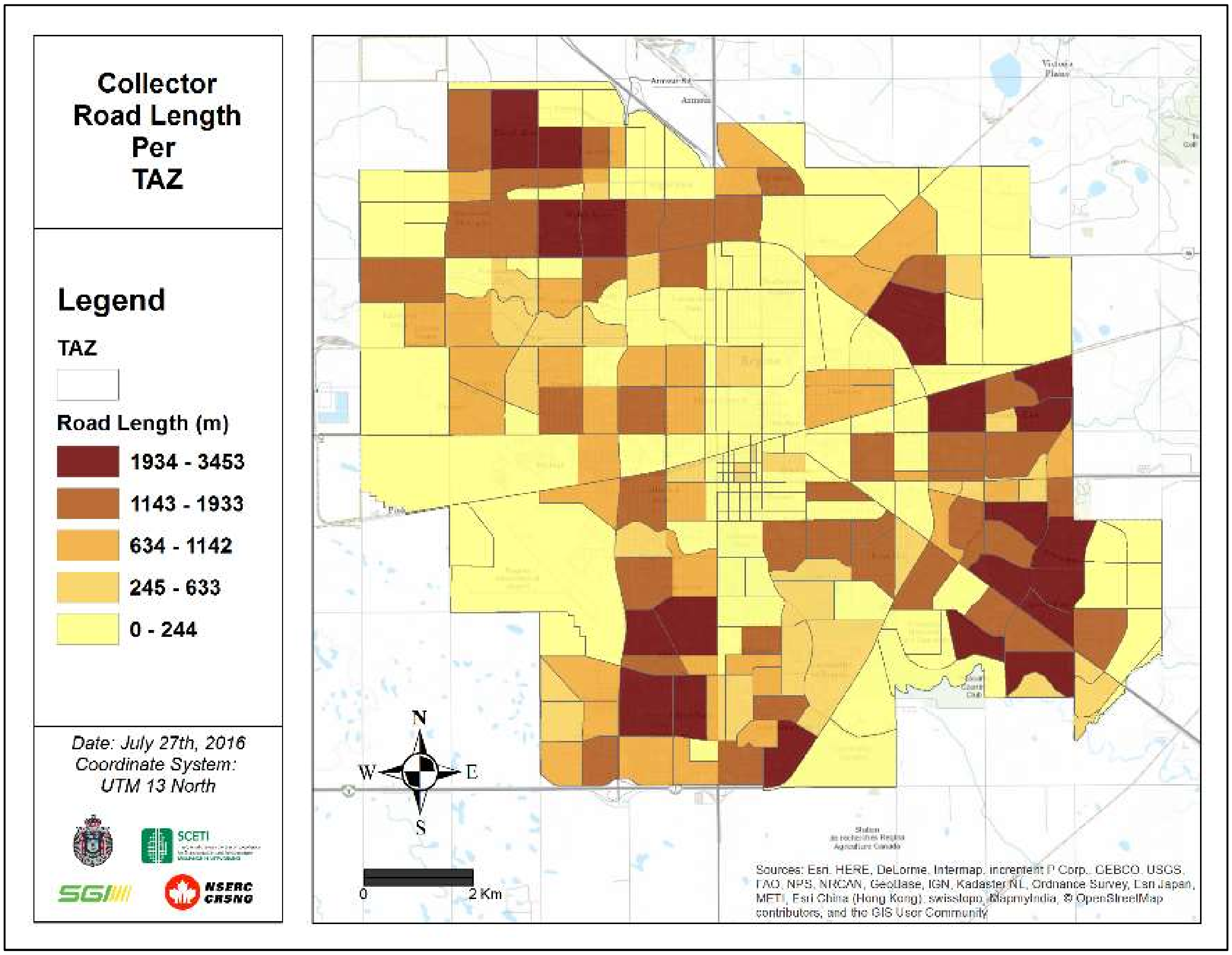
C4. Road Network and Infrastructure

Variable	Observations	Total	Min.	Maximum	Mean	Standard Deviation
Arterial Road Length	263	145654.95	0	2618.06	555.93	565.05
Collector Road Length (m)	263	162863.55	0	3452.70	621.16	751.80
Expressway Length (m)	263	22037	0	2029	84.11	262.64
Gravel Road Length (m)	263	35762	0	4879	136.50	488.31
Highway Length (m)	263	10810.67	0	1820.18	41.26	205.32
Local Road Length (m)	263	620715.03	0	9643.52	2369.14	2396.95
Private Road Length (m)	263	61802	0	12726	235.89	995.59
Ramp Length (m)	263	23831.47	0	1825.64	90.96	275.05
Right-Of-Way Length (m)	263	1074	0	434	4.10	32.73
Total Road Segment Length (m)	263	1084721.83	87.00	16354.95	4140.16	3299.99
Average Road Segment Length (m)	263	563389.83	81.17	11361.34	2150.34	1828.17
Roadway Length with Average Speed Limit (m)	263	22277.33	40	536.17	85.03	55.84
Road Segment Length with posted Speed Limit 20km/hr	263	200	0	200	0.76	12.33
Road Segment Length with posted Speed Limit 30km/hr	263	1179	0	523	4.5	45.01
Road Segment Length with posted Speed Limit 40km/hr	263	166654.74	0	3836.75	636.09	818.10
Road Segment Length with posted Speed Limit 50km/hr	263	824639.38	0	14405.84	3147.48	2703.98
Road Segment Length with posted Speed Limit 60km/hr	263	7485	0	1695	28.57	151.33
Road Segment Length with posted Speed Limit 70km/hr	263	32074.57	0	3284.75	122.42	392.76
Road Segment Length with posted Speed Limit 80km/hr	263	34125	0	3677	130.25	431.64
Road Segment Length with posted Speed Limit 100km/hr	263	18366	0	1820	70.10	251.46
Number of three-leg intersections	263	2362	0	66	9.02	10.02
Number of four-leg intersections	263	1364	0	36	5.21	6.41
Number of five-leg intersections	263	4	0	1	0.02	0.12
Total Number of Intersections	263	3725	0	79	14.22	13.63
Vehicle-Kilometer-Traveled (VKMT)	263	6261583.42	43.43	122800.88	23899.17	21490.66
Annual Average Daily Traffic (AADT)	263	32628715.58	700	537742.28	124537.08	101783.47
Total Lane Kilometer Traveled, TLKM (km)	263	1084.72	0.087	16.35	4.14	3.30
Intersection Density, INTKD (Number of intersection per TLKM)	263	836.32	0	18.35	3.19	1.79
Proportion of three-leg intersections per TAZ Area (I3WP)	263	14248.66	0	100	54.38	34.22
ALKP	263	5191.54	0	100	19.82	23.85
LLKP	263	12413.35	0	100	47.38	30.64

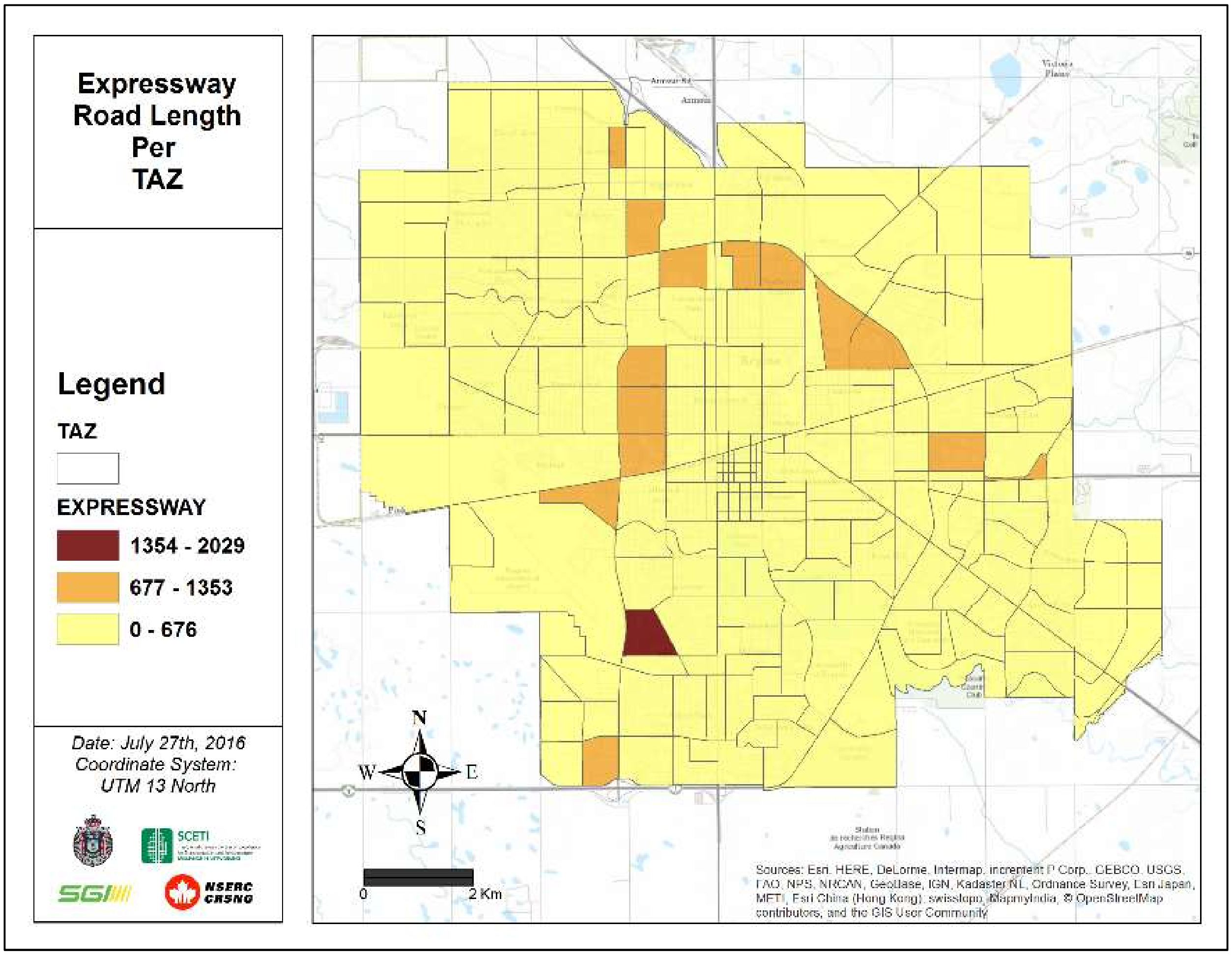
Arterial Road Length Per TAZ (m)



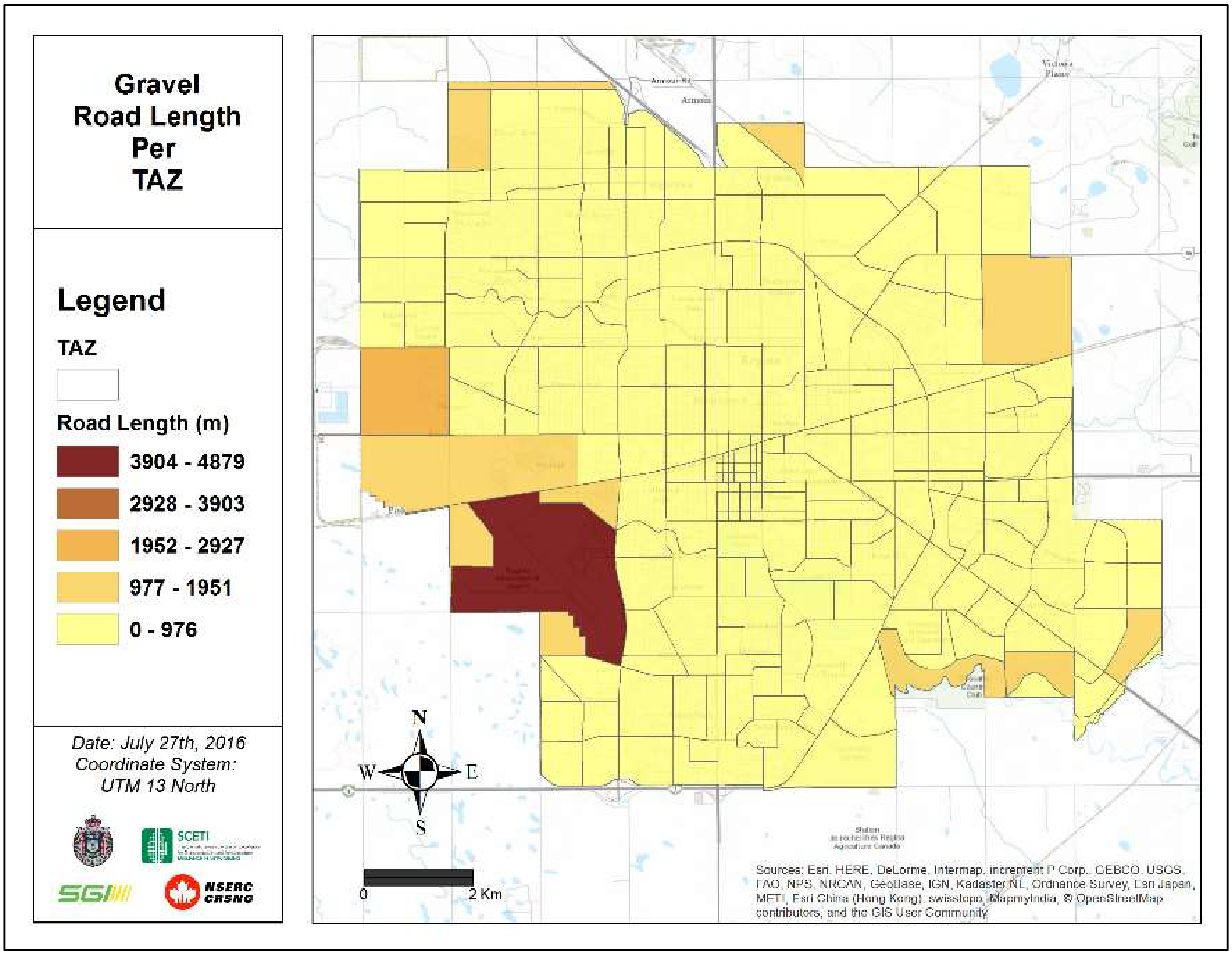
Collector Road Length Per TAZ (m)



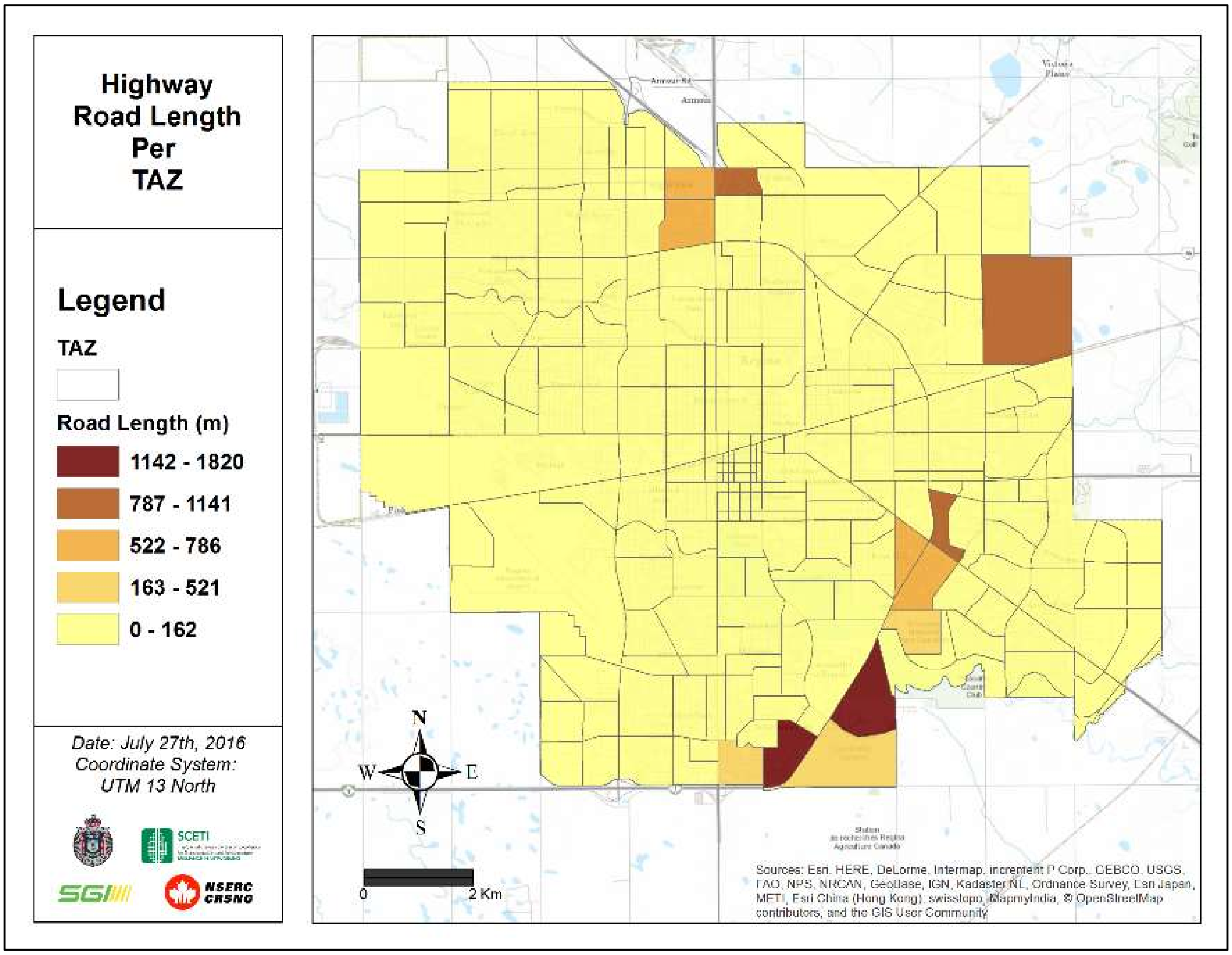
Expressway Road Length Per TAZ (m)



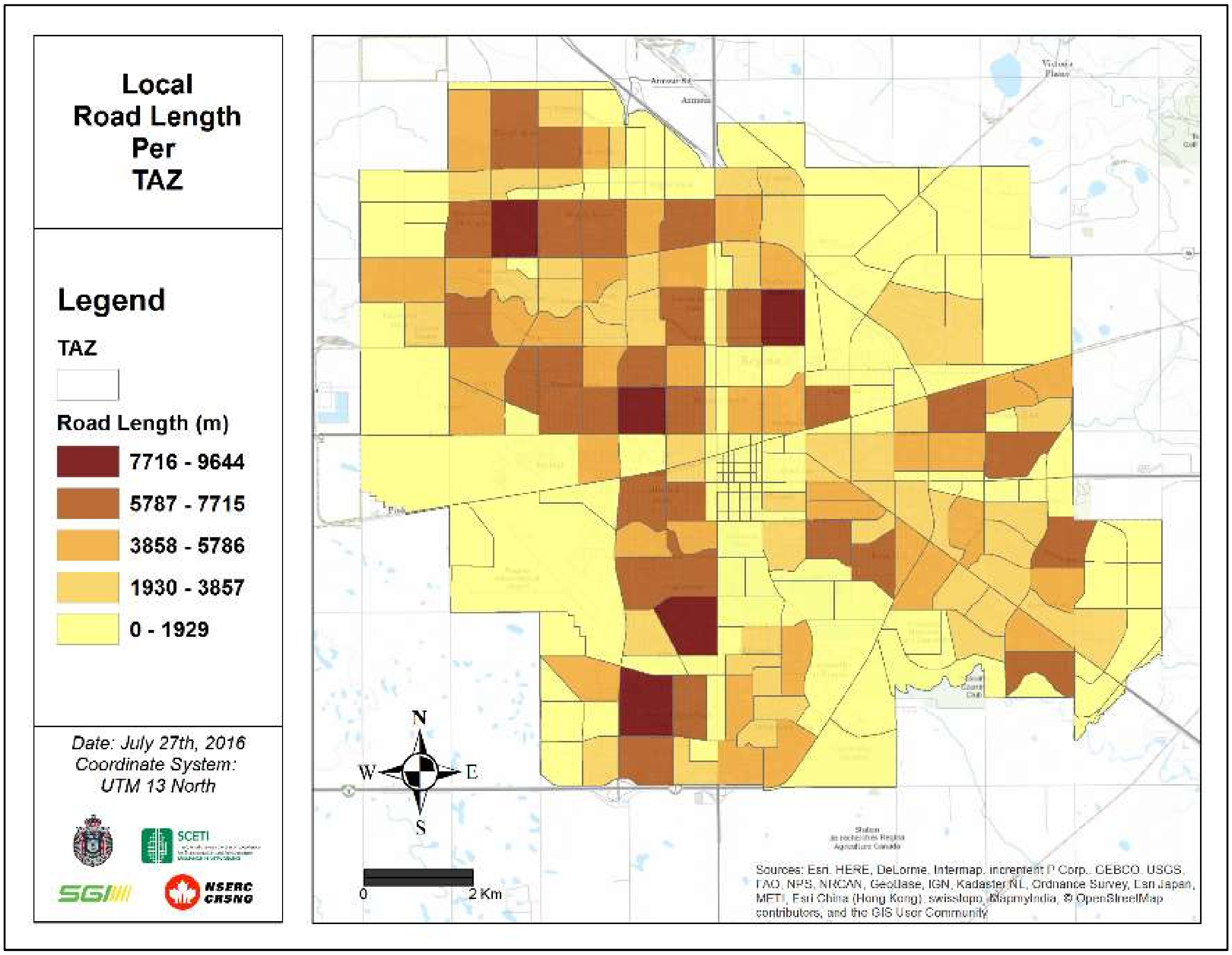
Gravel Road Length Per TAZ (m)



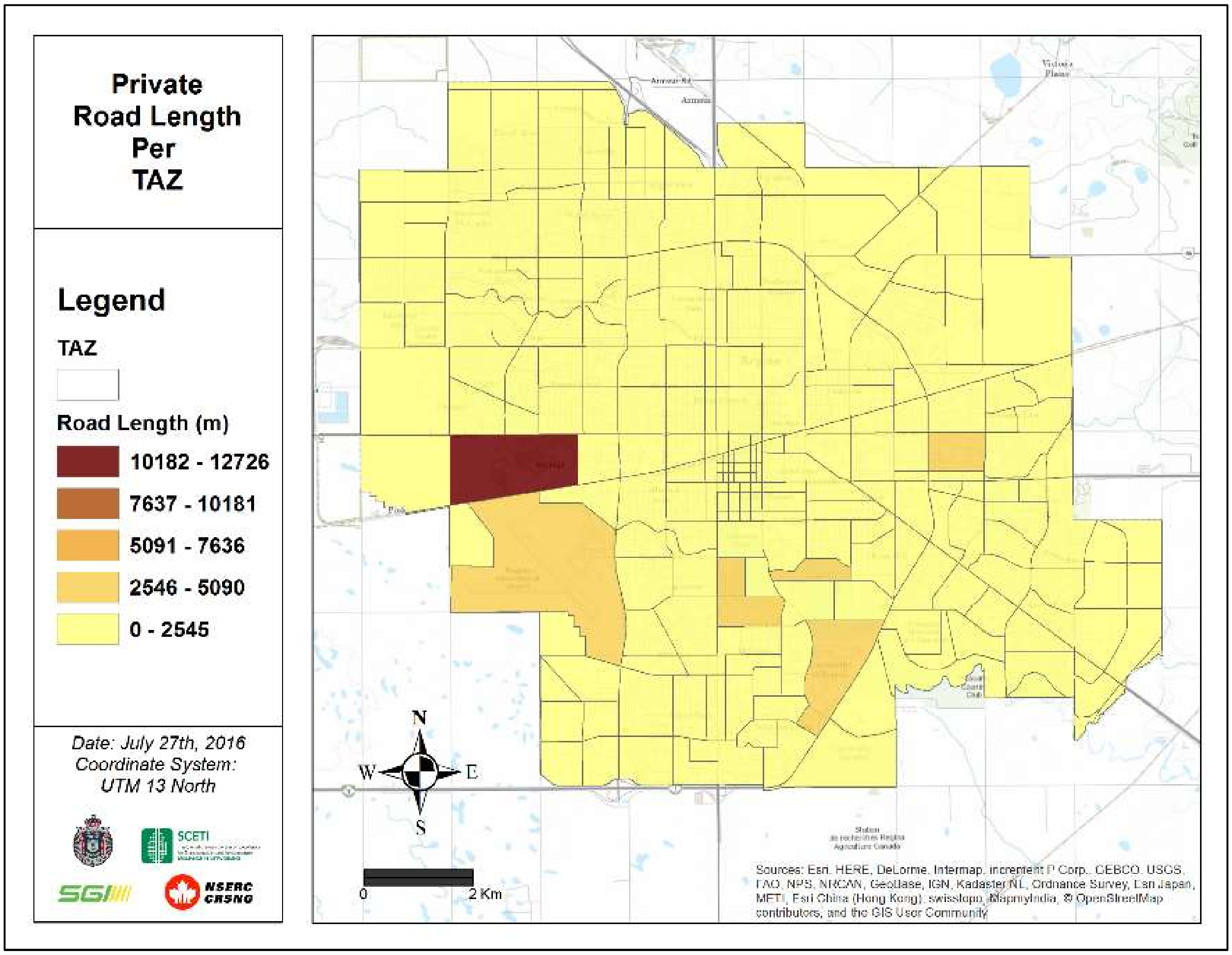
Highway Length Per TAZ (m)



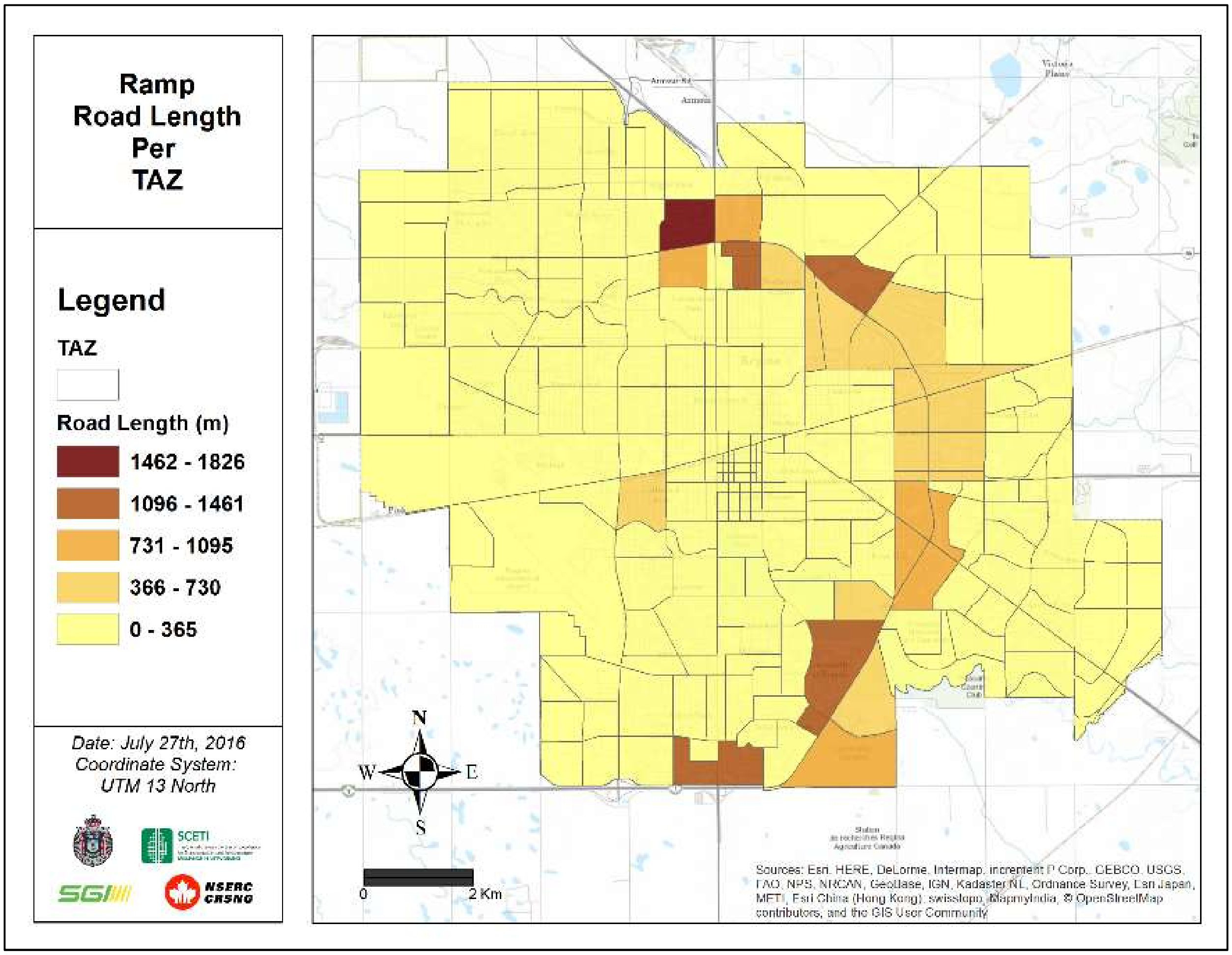
Local Road Length Per TAZ (m)



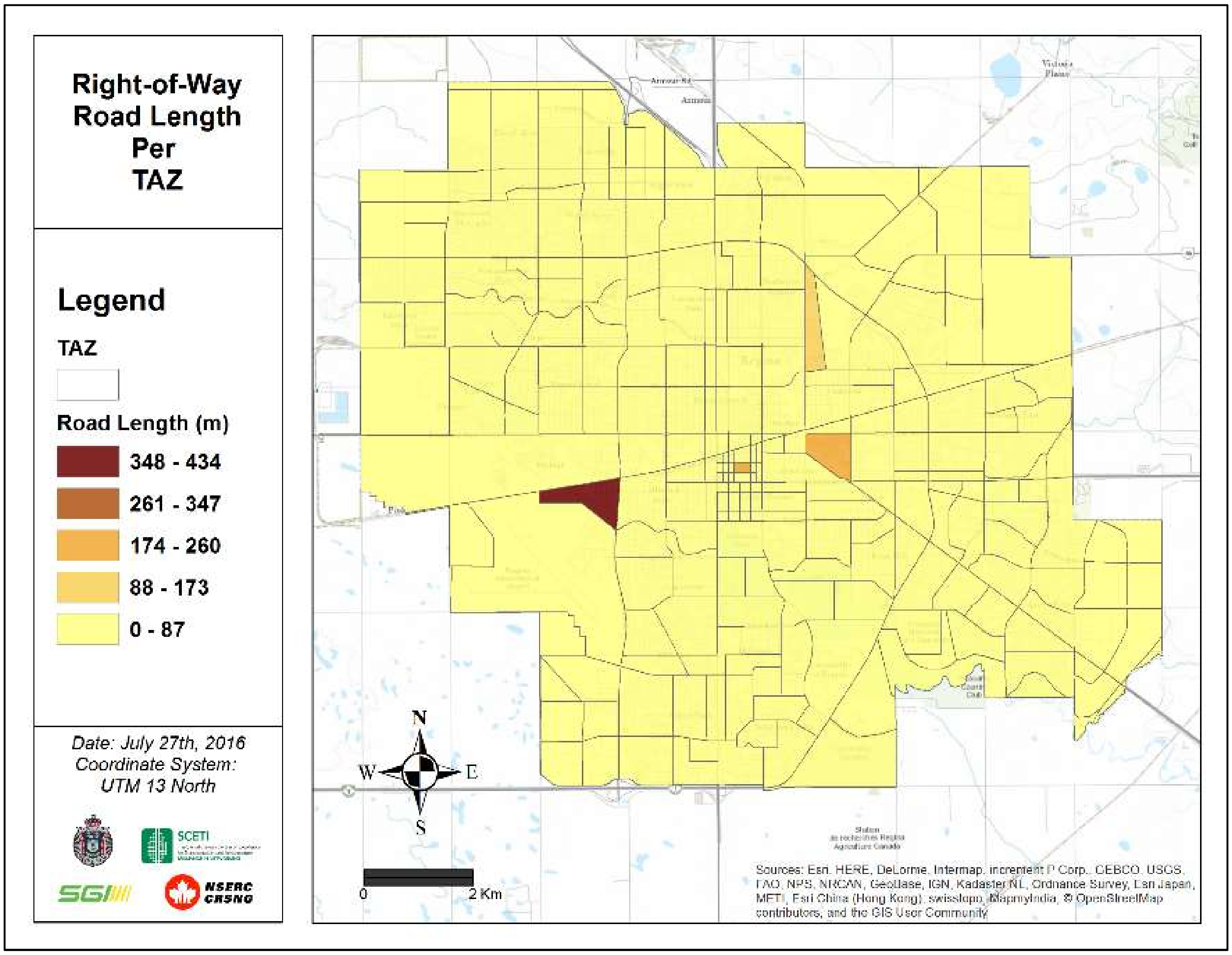
Private Road Length Per TAZ (m)



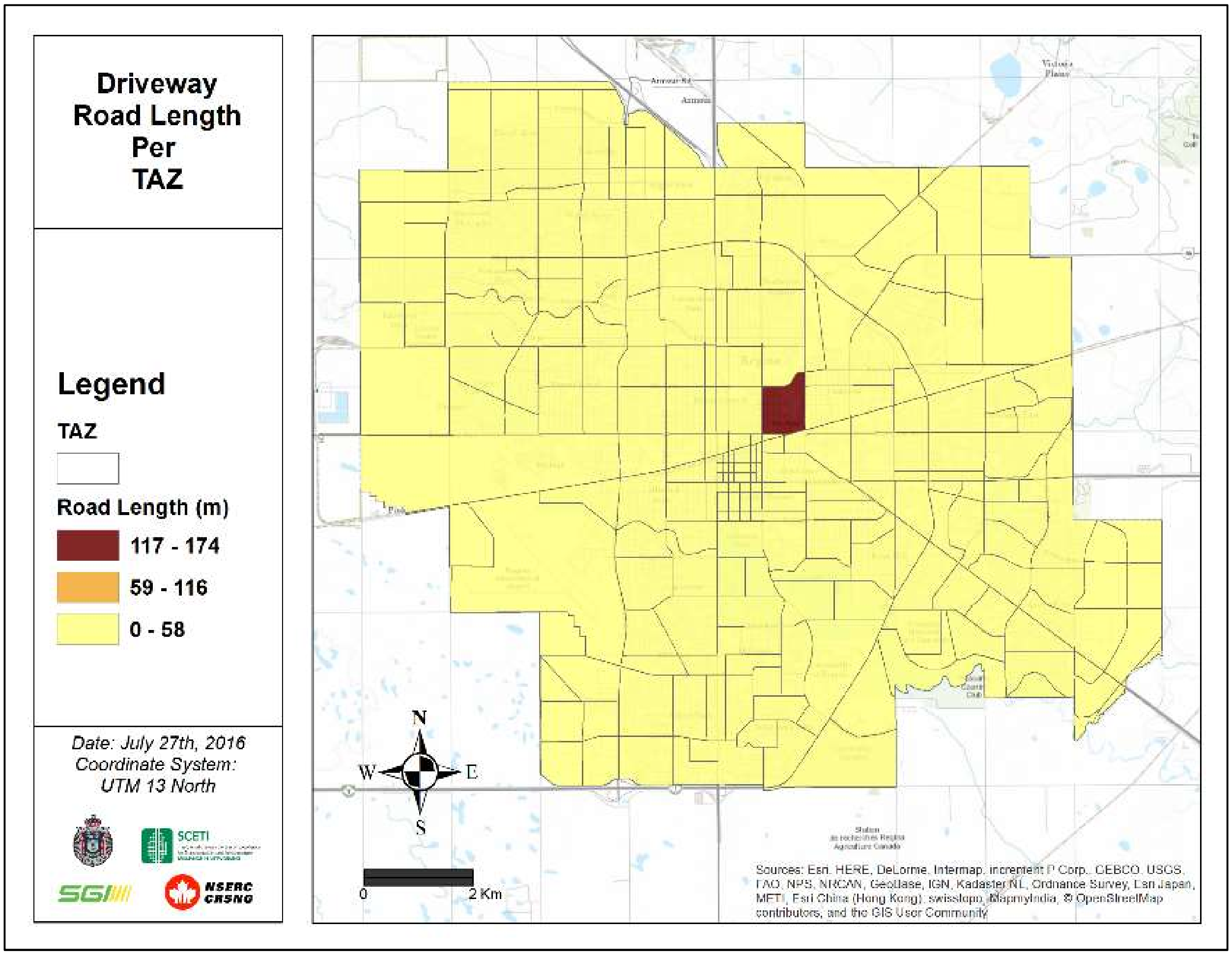
Ramp Length Per TAZ (m)



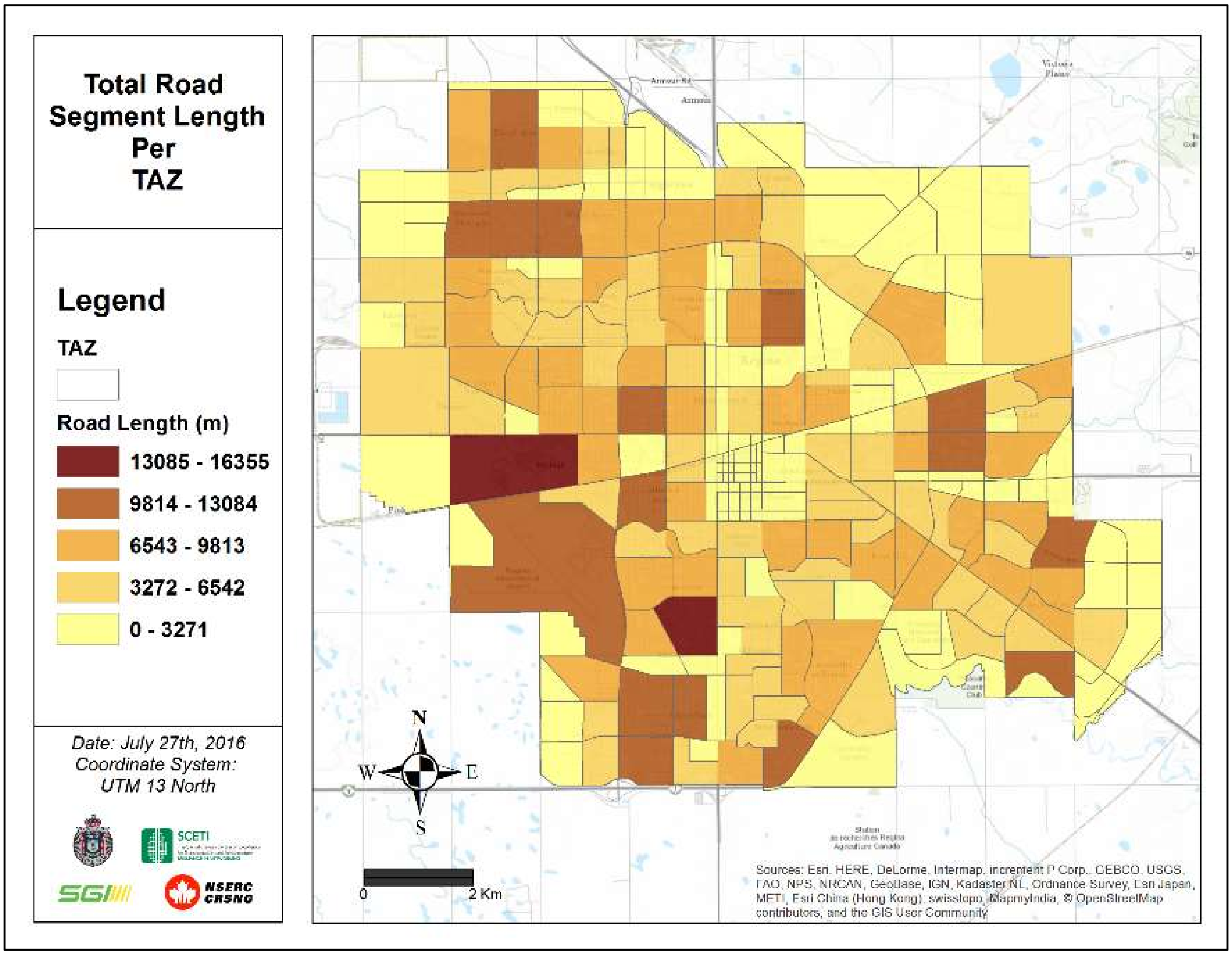
Right-of-Way Road Length Per TAZ (m)



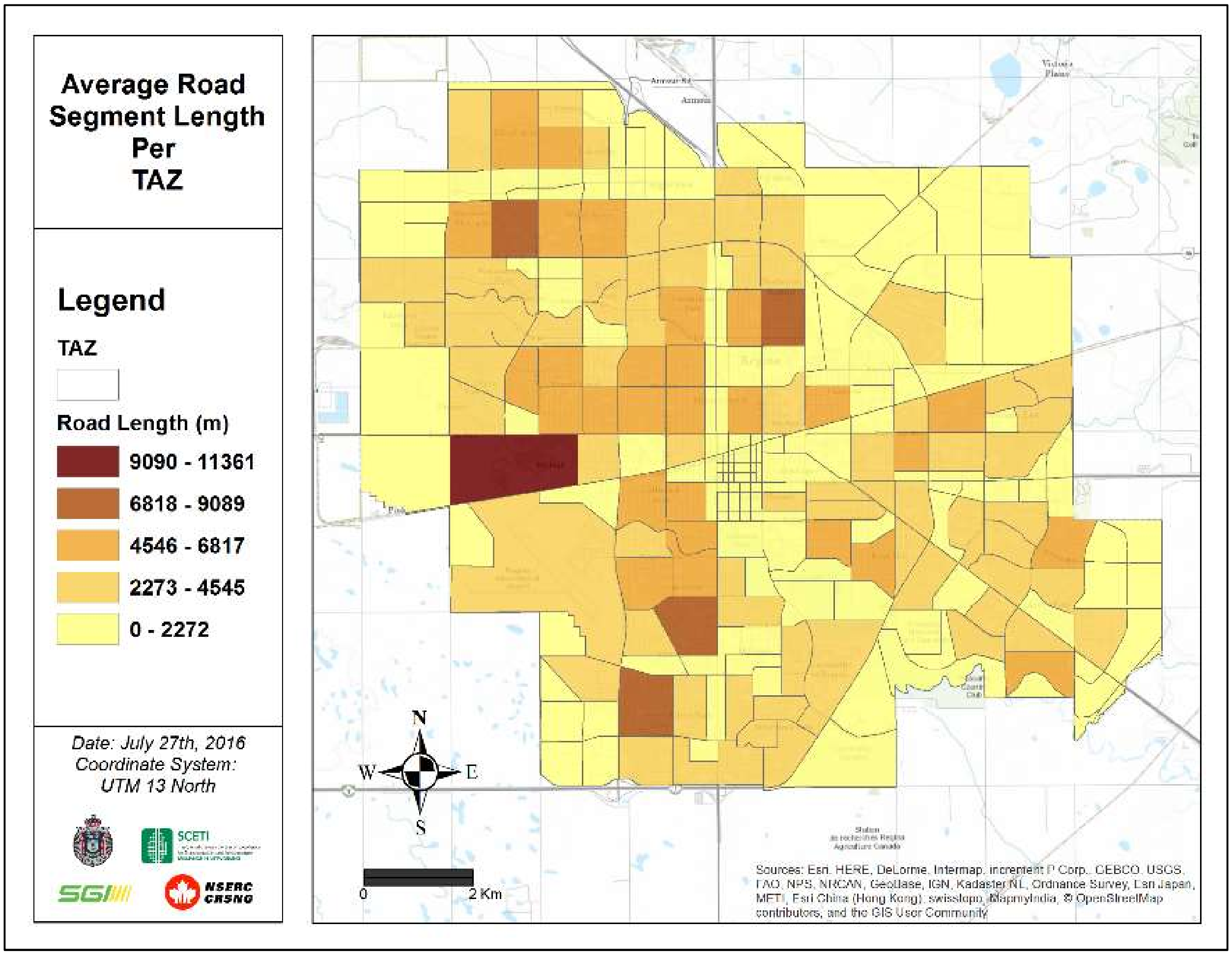
Driveway Road Length Per TAZ (m)



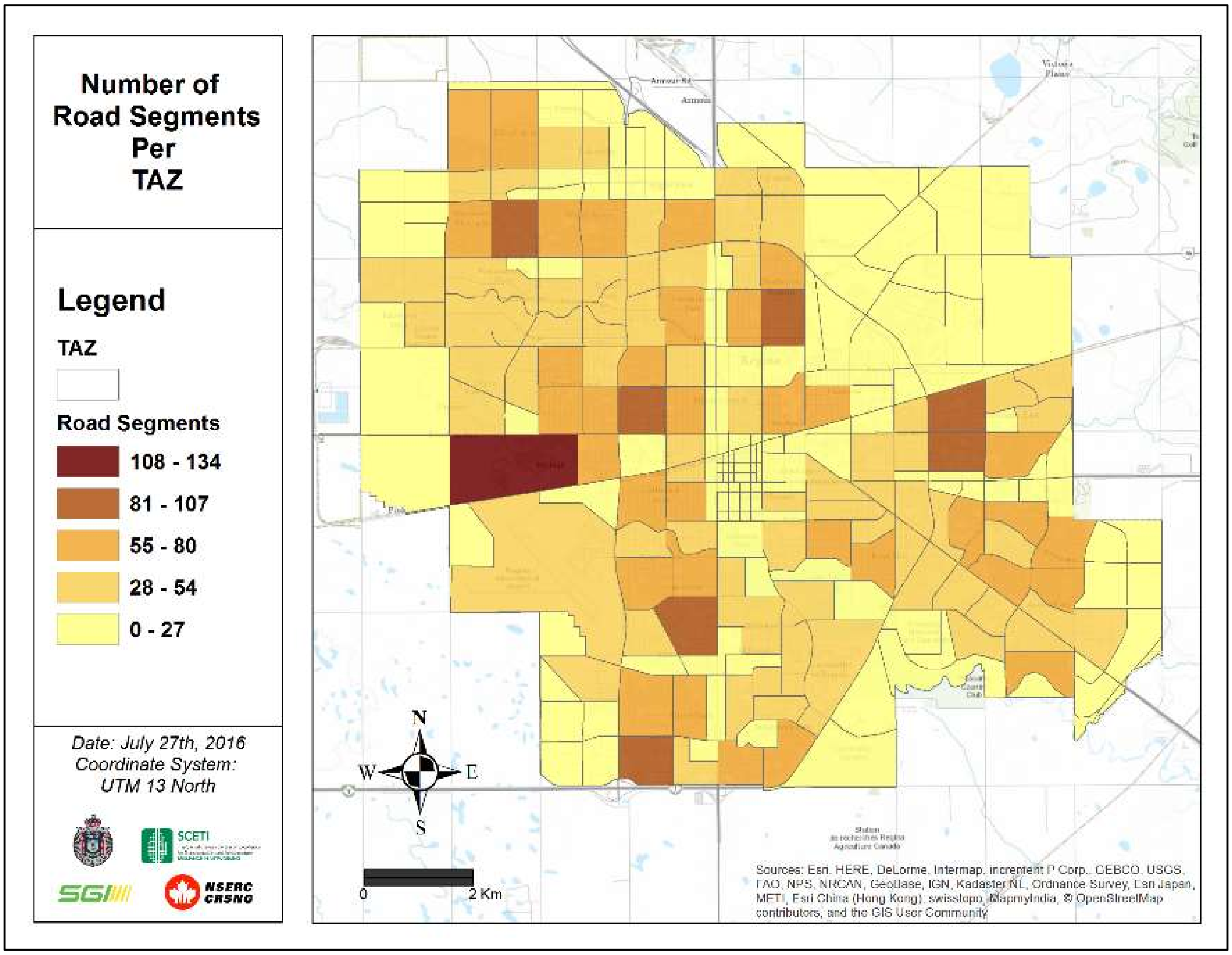
Total Road Segment Length Per TAZ (m)



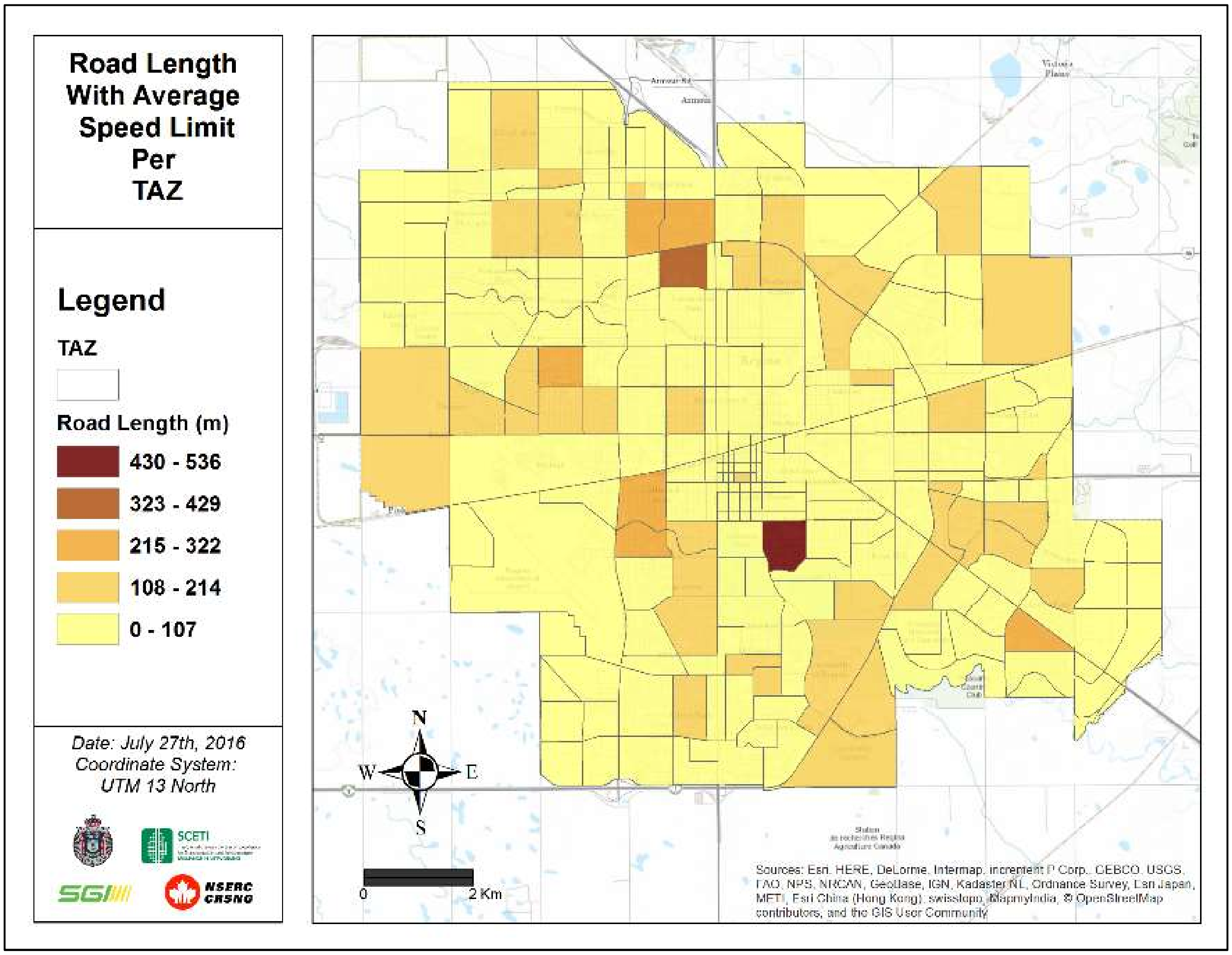
Average Road Segment Length Per TAZ (m)



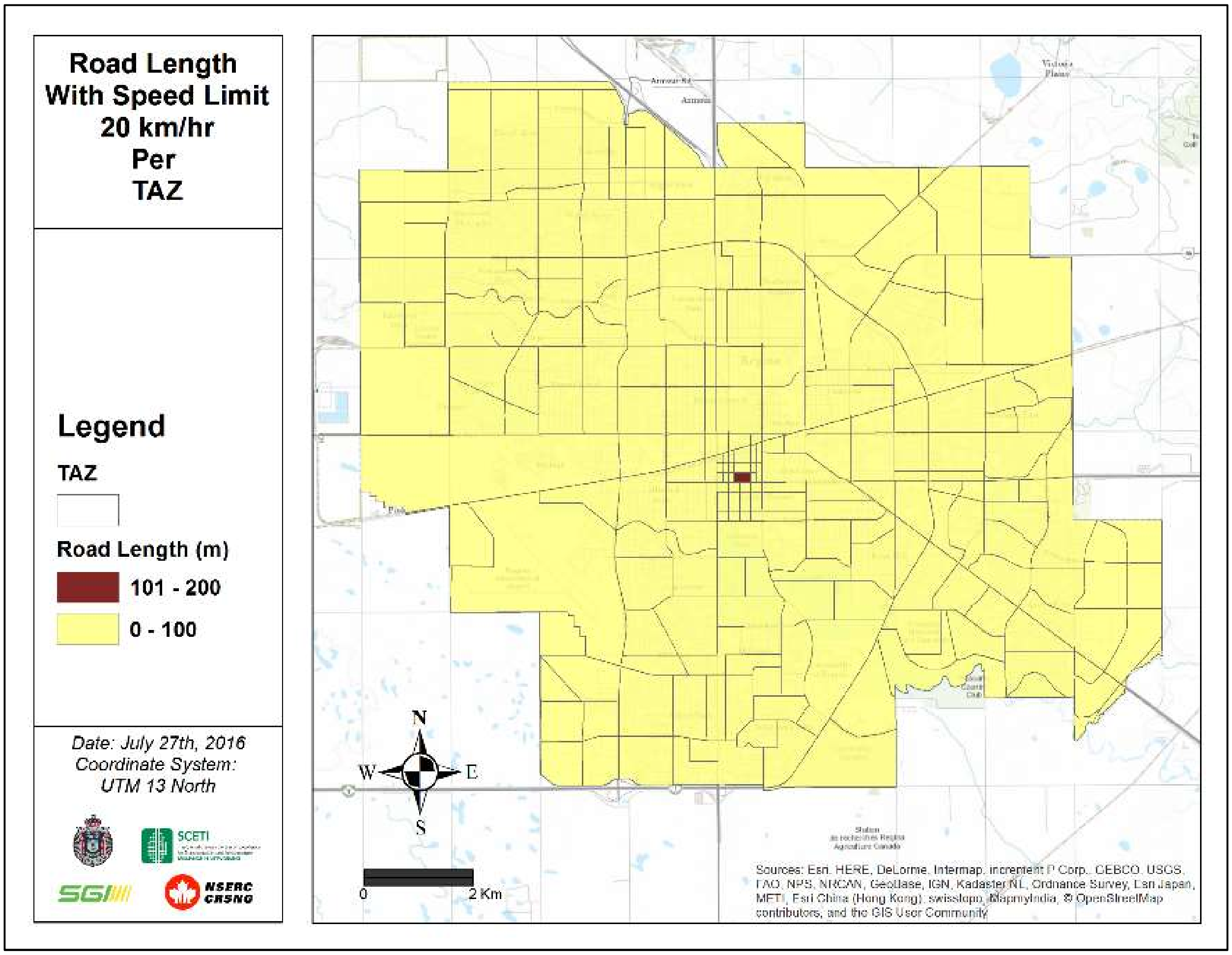
Number of Road Segment Length Per TAZ



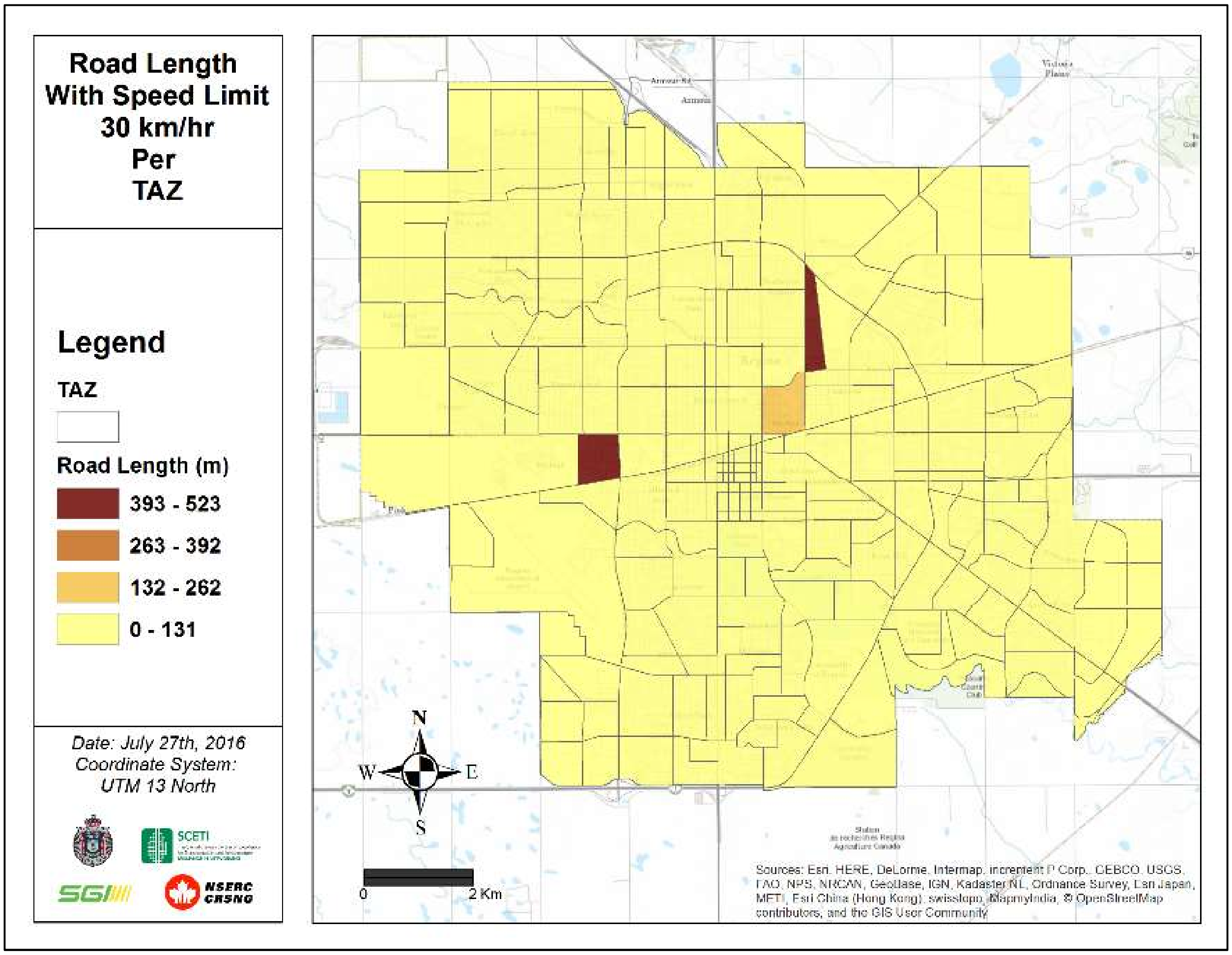
Road Segment Length with Average Speed Limit Per TAZ (m)



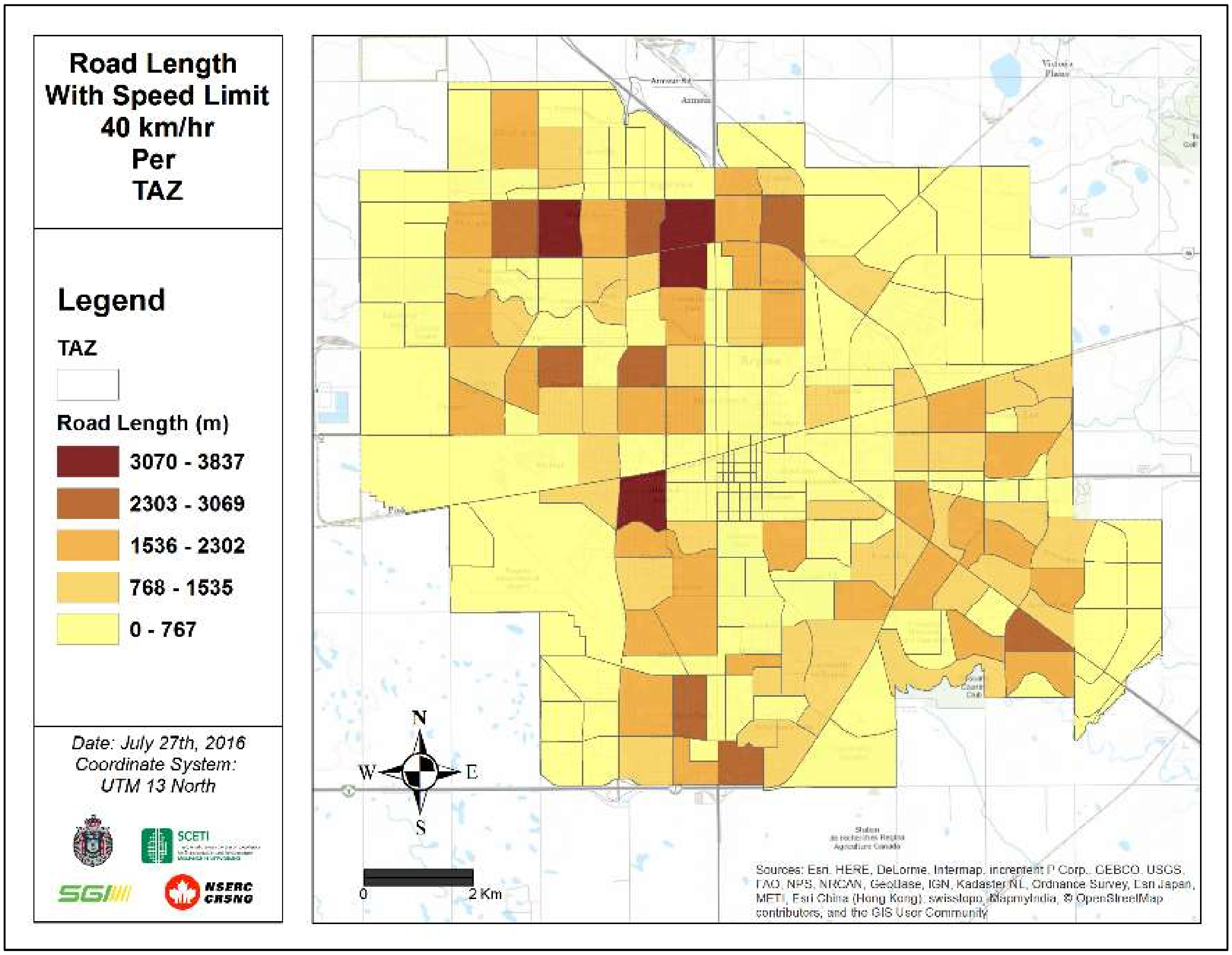
Road Segment Length with Posted Speed Limit of 20km/hr Per TAZ (m)



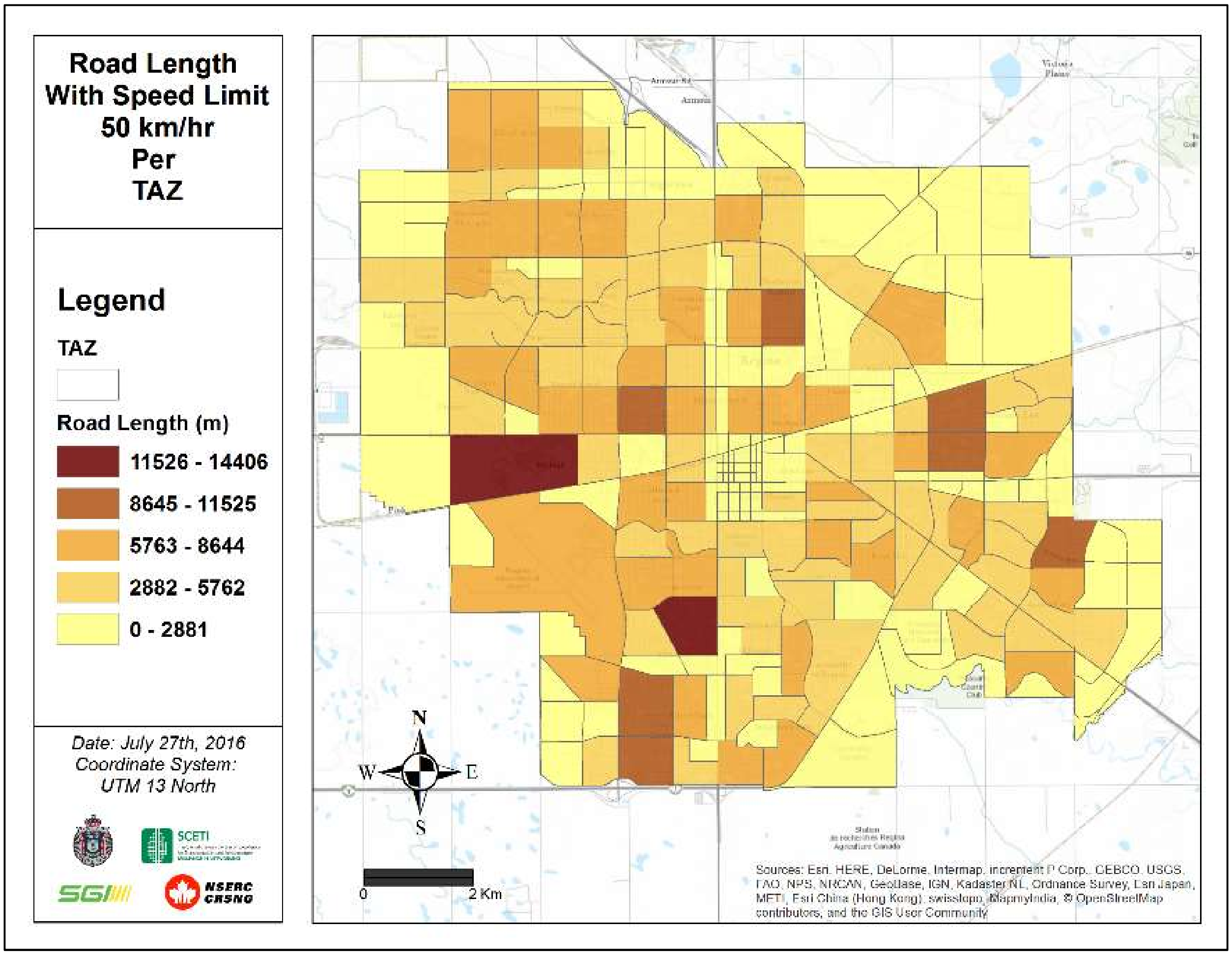
Road Segment Length with Posted Speed Limit of 30km/hr Per TAZ (m)



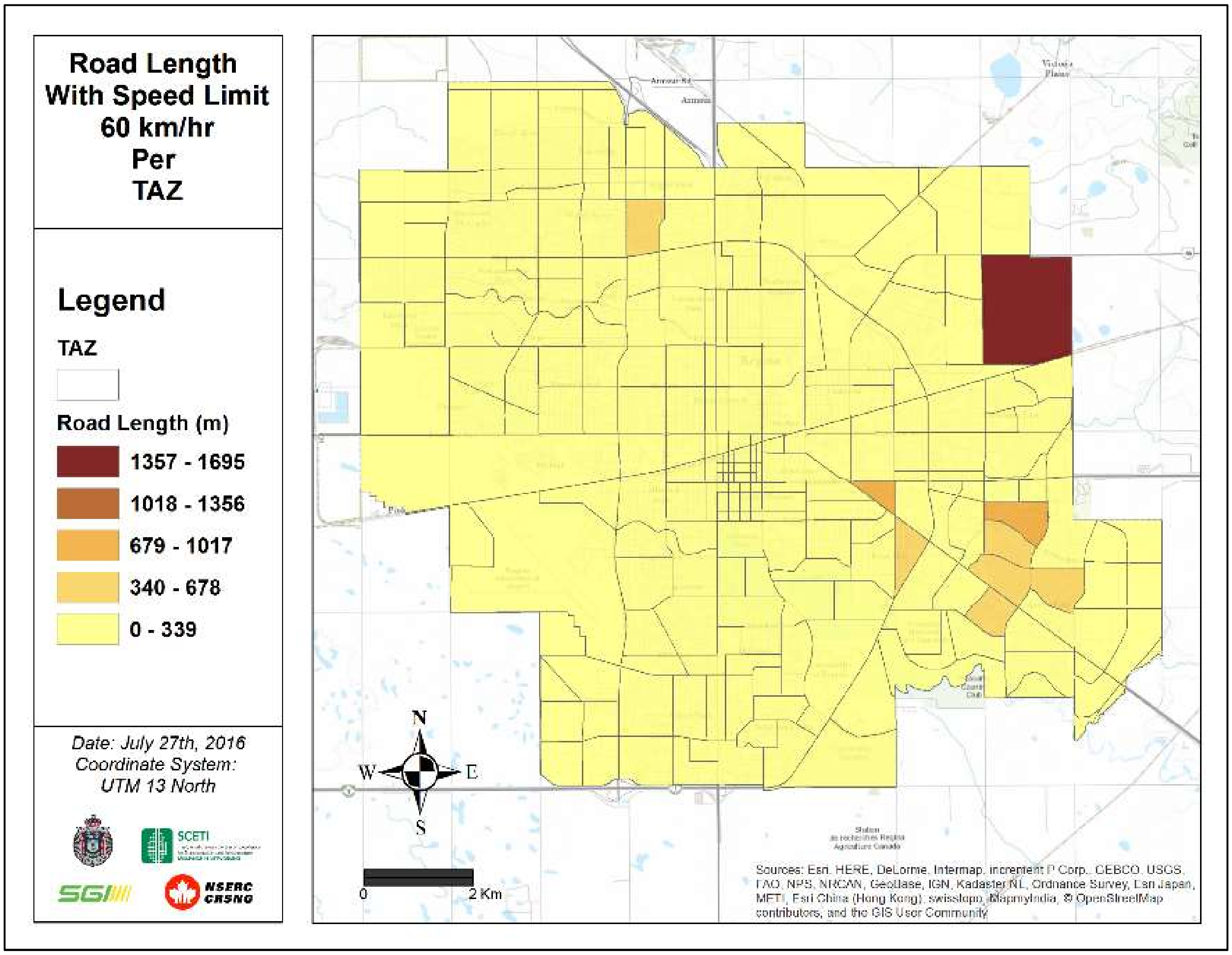
Road Segment Length with Posted Speed Limit of 40km/hr Per TAZ (m)



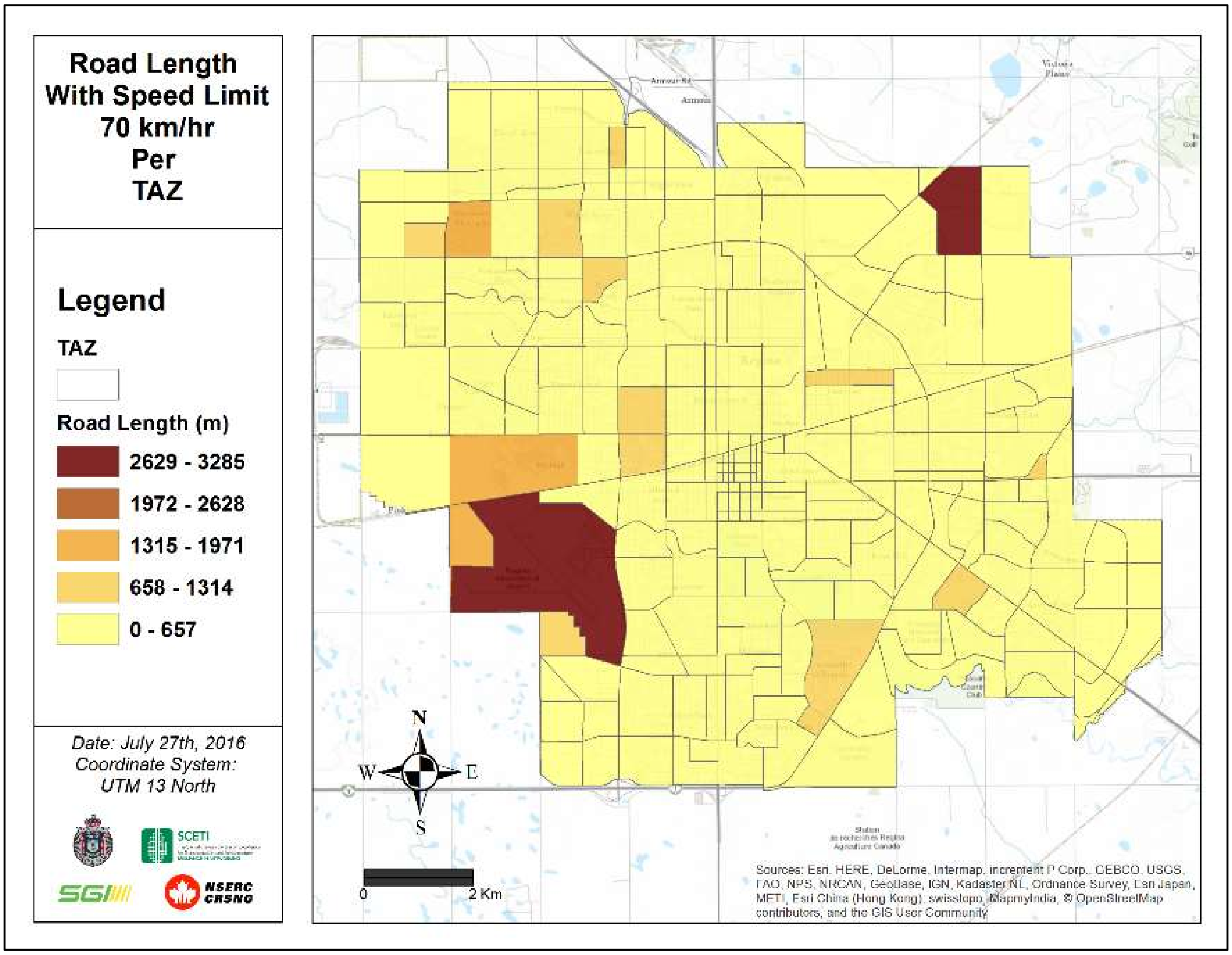
Road Segment Length with Posted Speed Limit of 50km/hr Per TAZ (m)



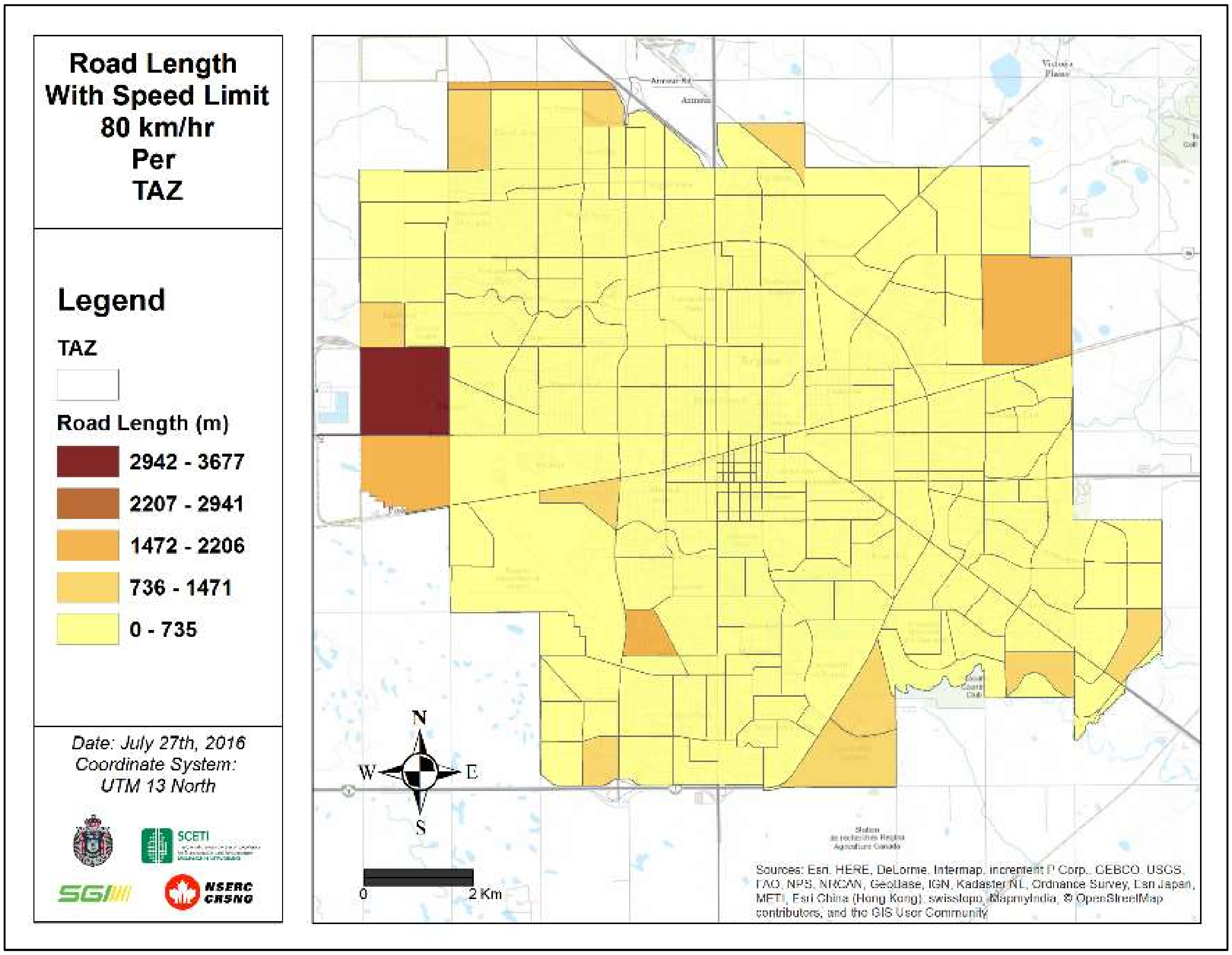
Road Segment Length with Posted Speed Limit of 60km/hr Per TAZ (m)



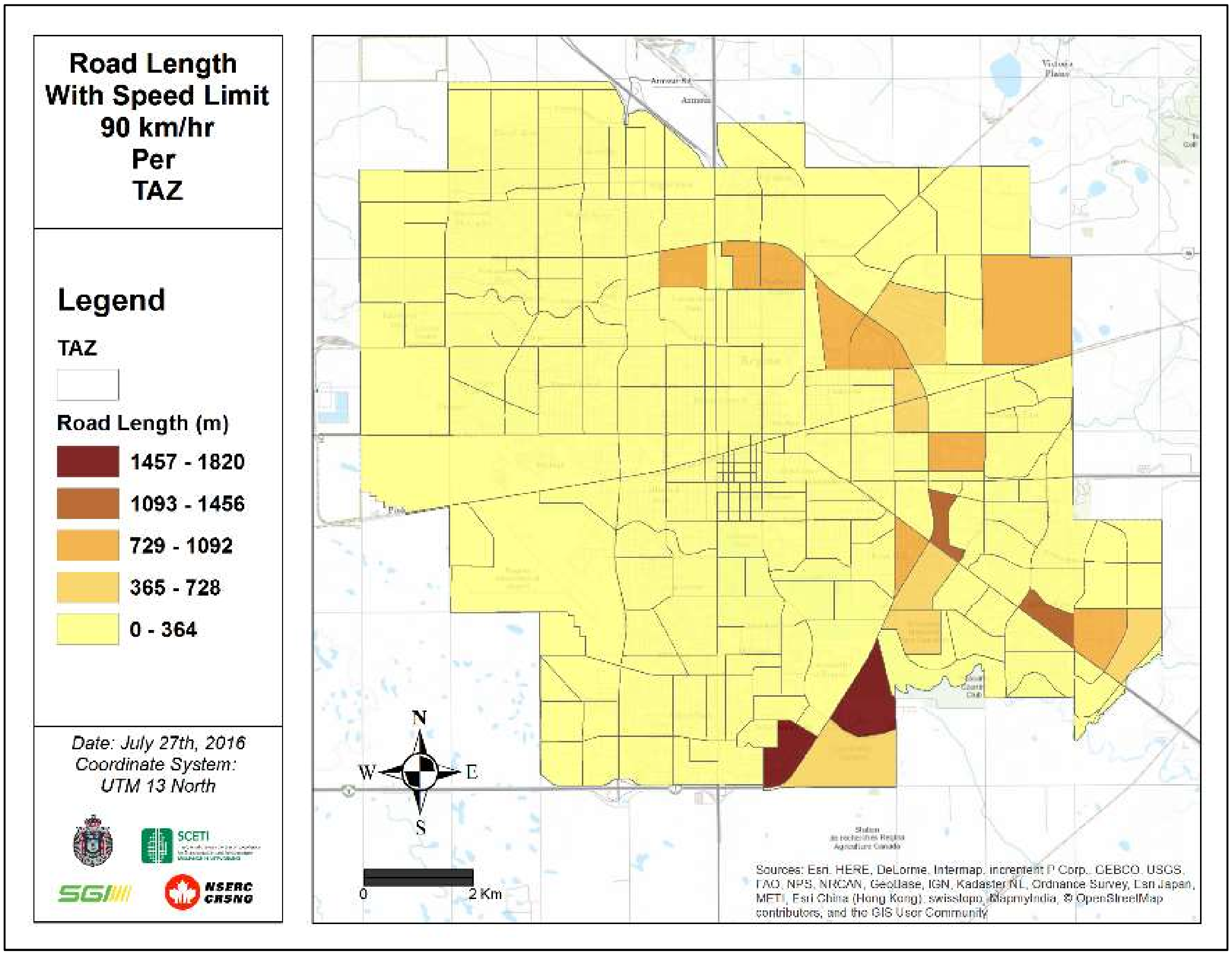
Road Segment Length with Posted Speed Limit of 70km/hr Per TAZ (m)



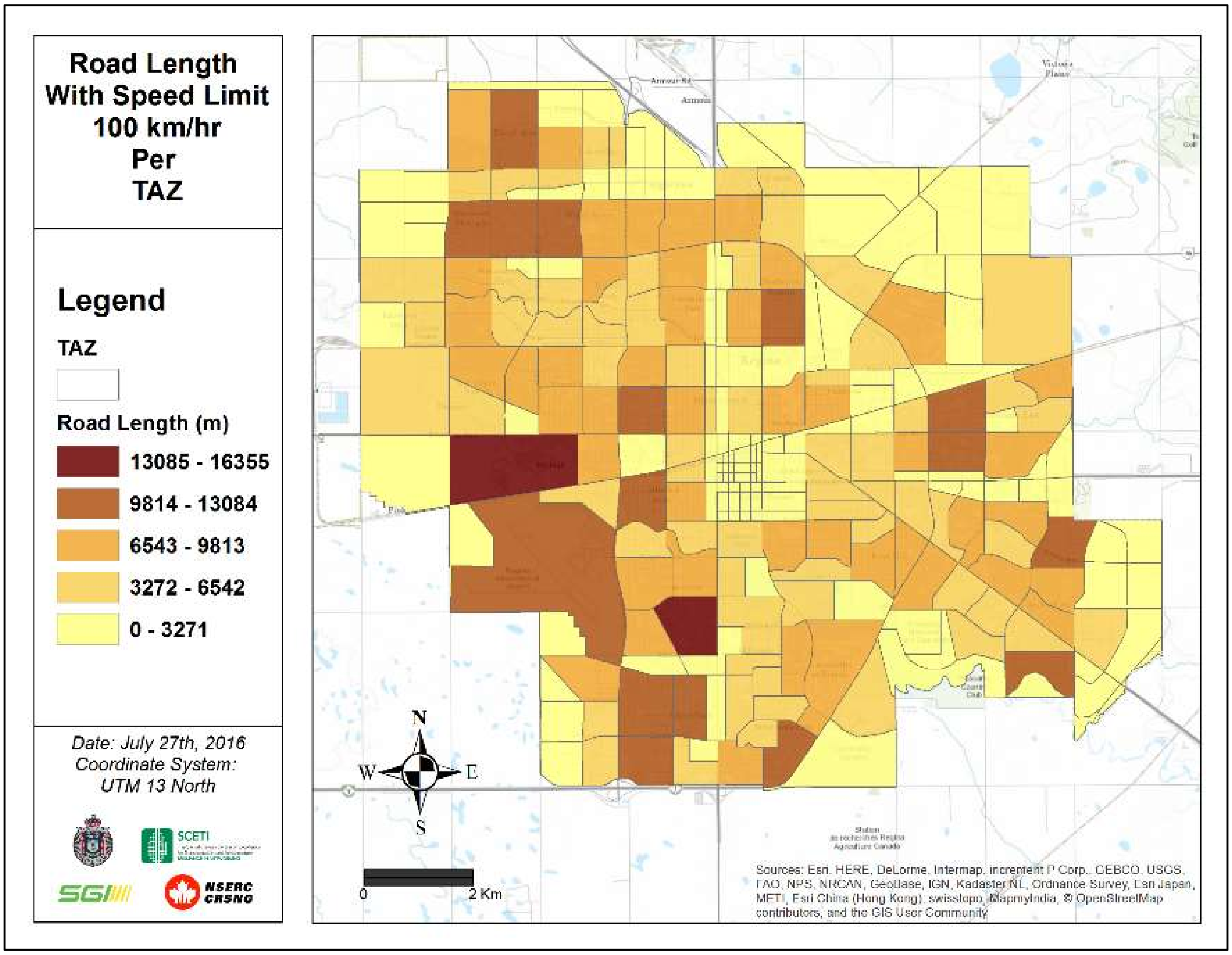
Road Segment Length with Posted Speed Limit of 80km/hr Per TAZ (m)



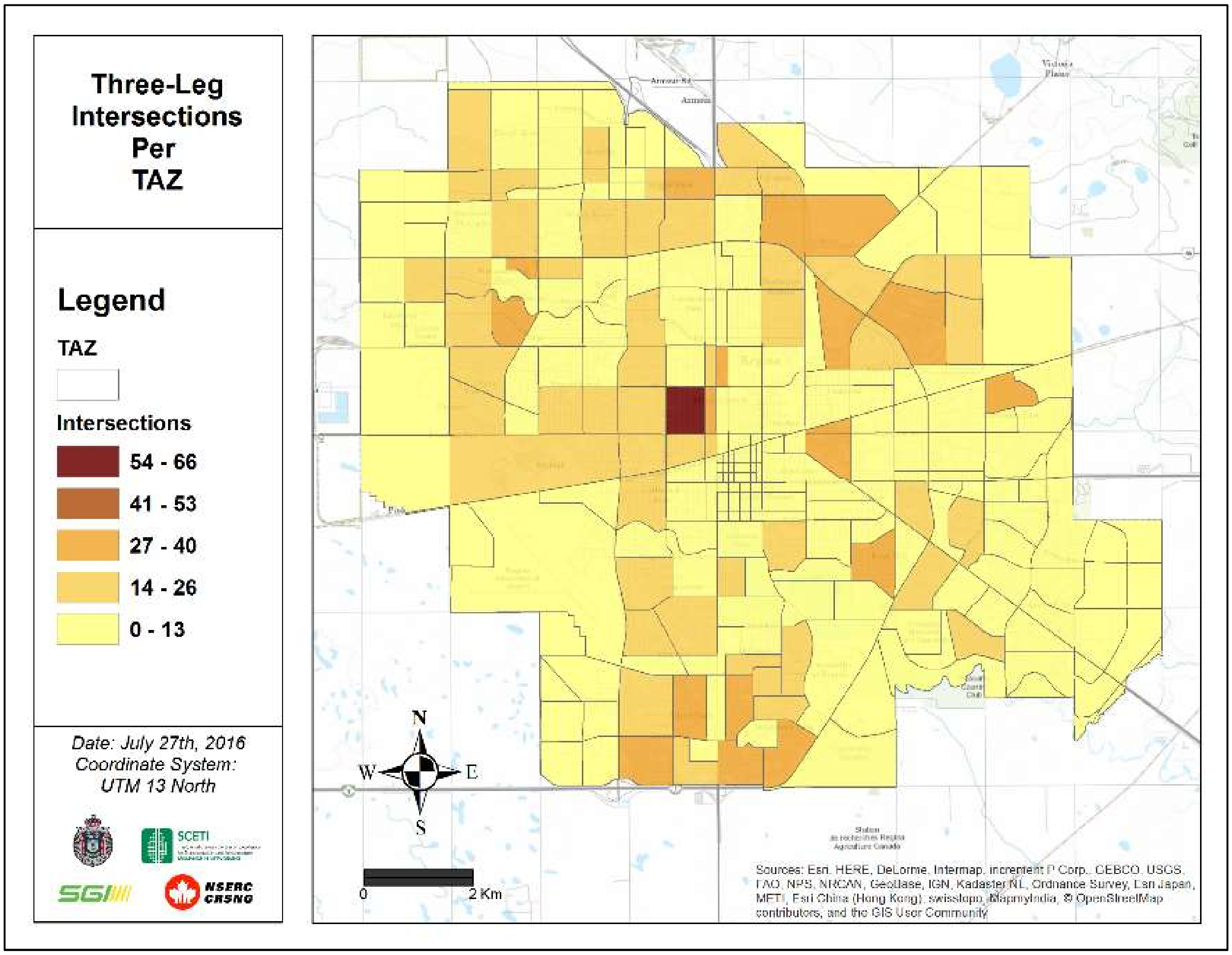
Road Segment Length with Posted Speed Limit of 90km/hr Per TAZ (m)



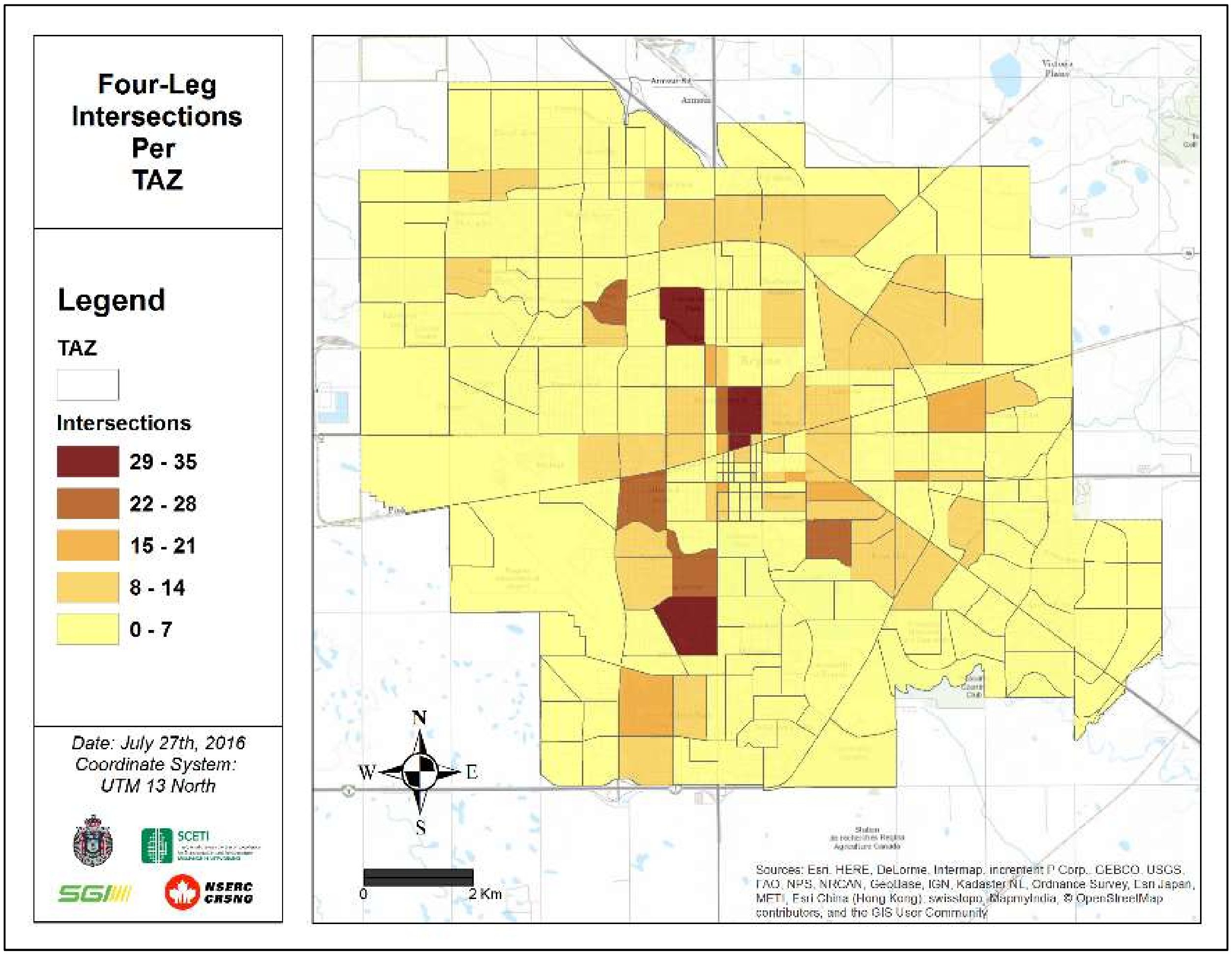
Road Segment Length with Posted Speed Limit of 100km/hr Per TAZ (m)



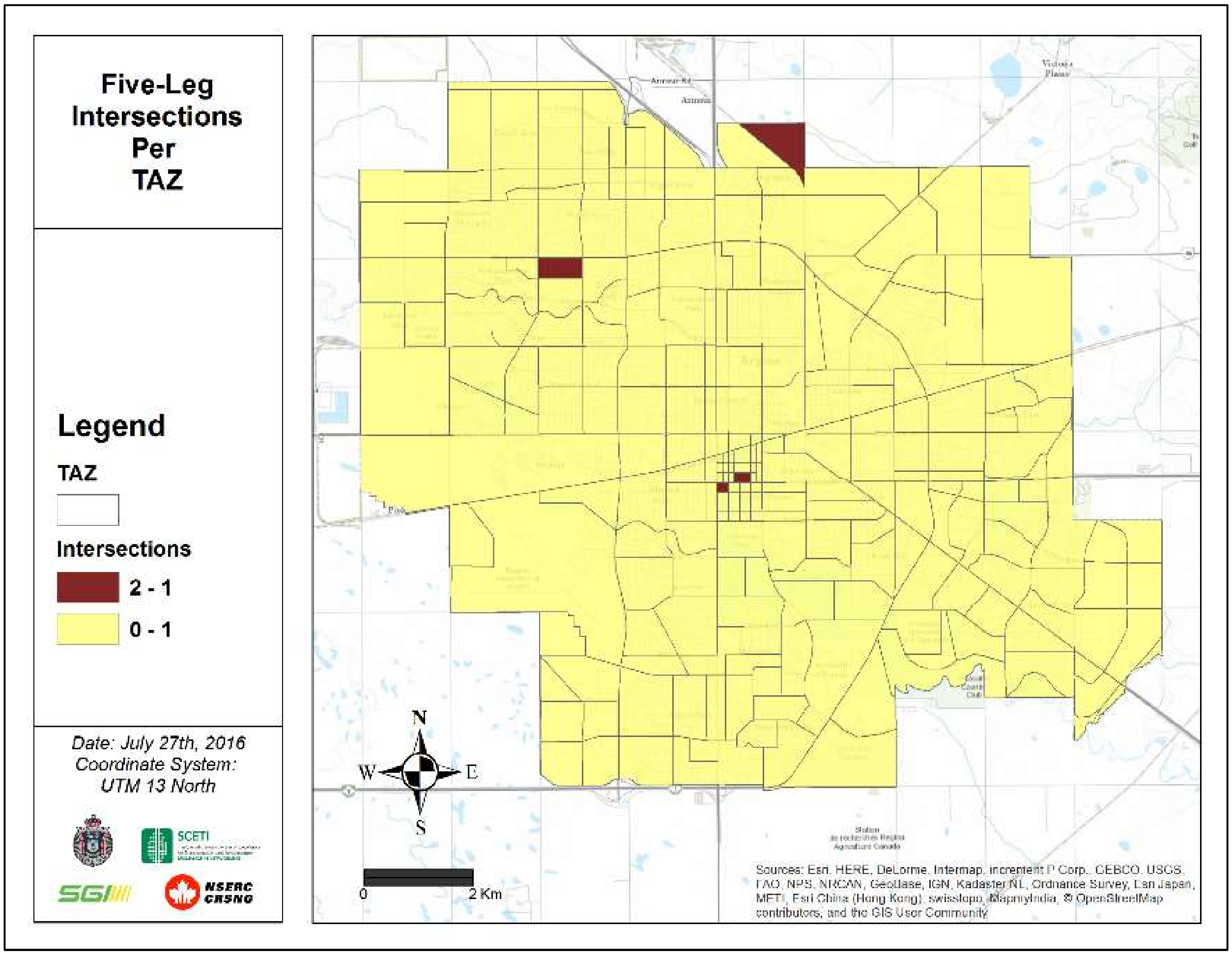
Number of Three-Leg Intersections Per TAZ



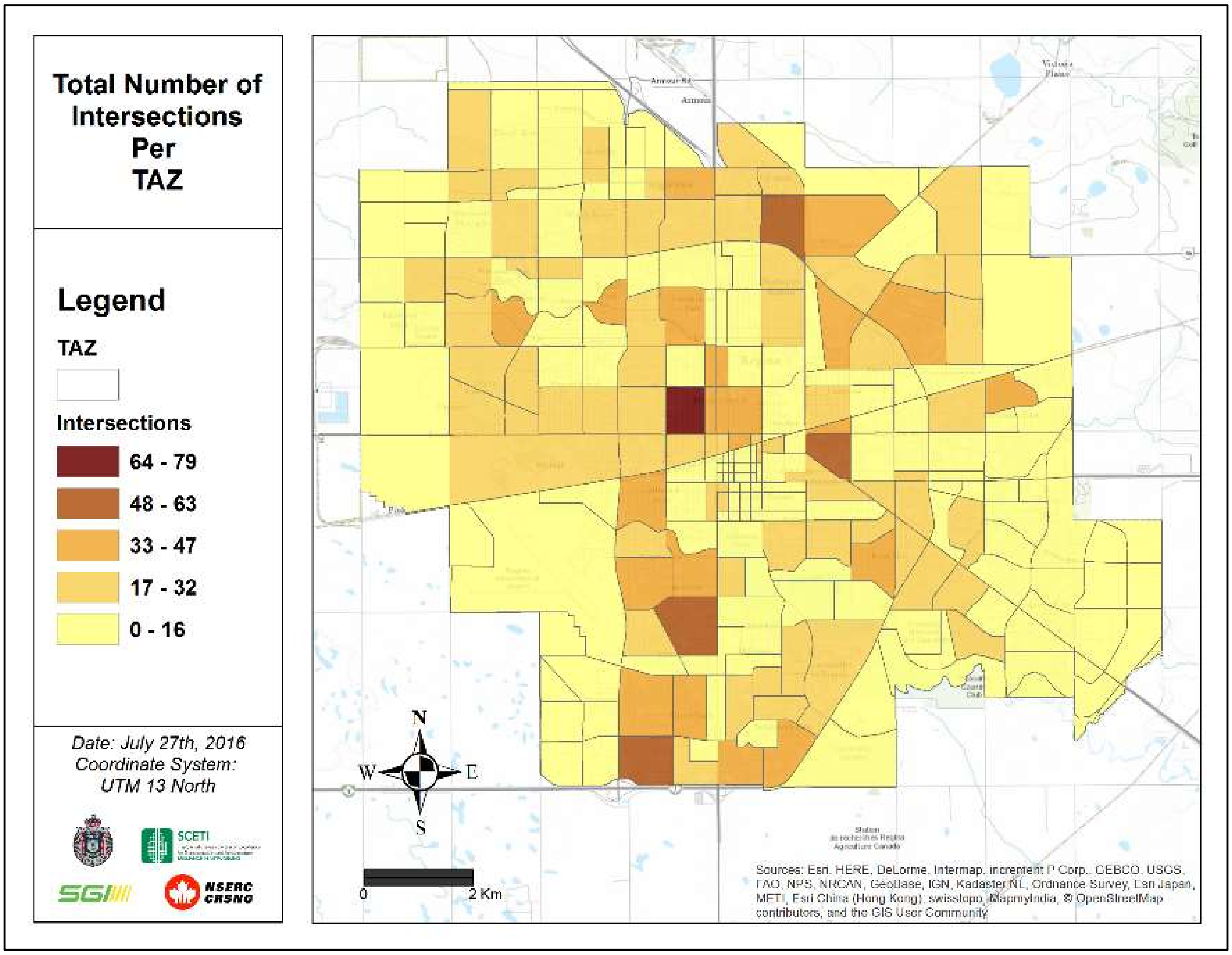
Number of Four-Leg Intersections Per TAZ



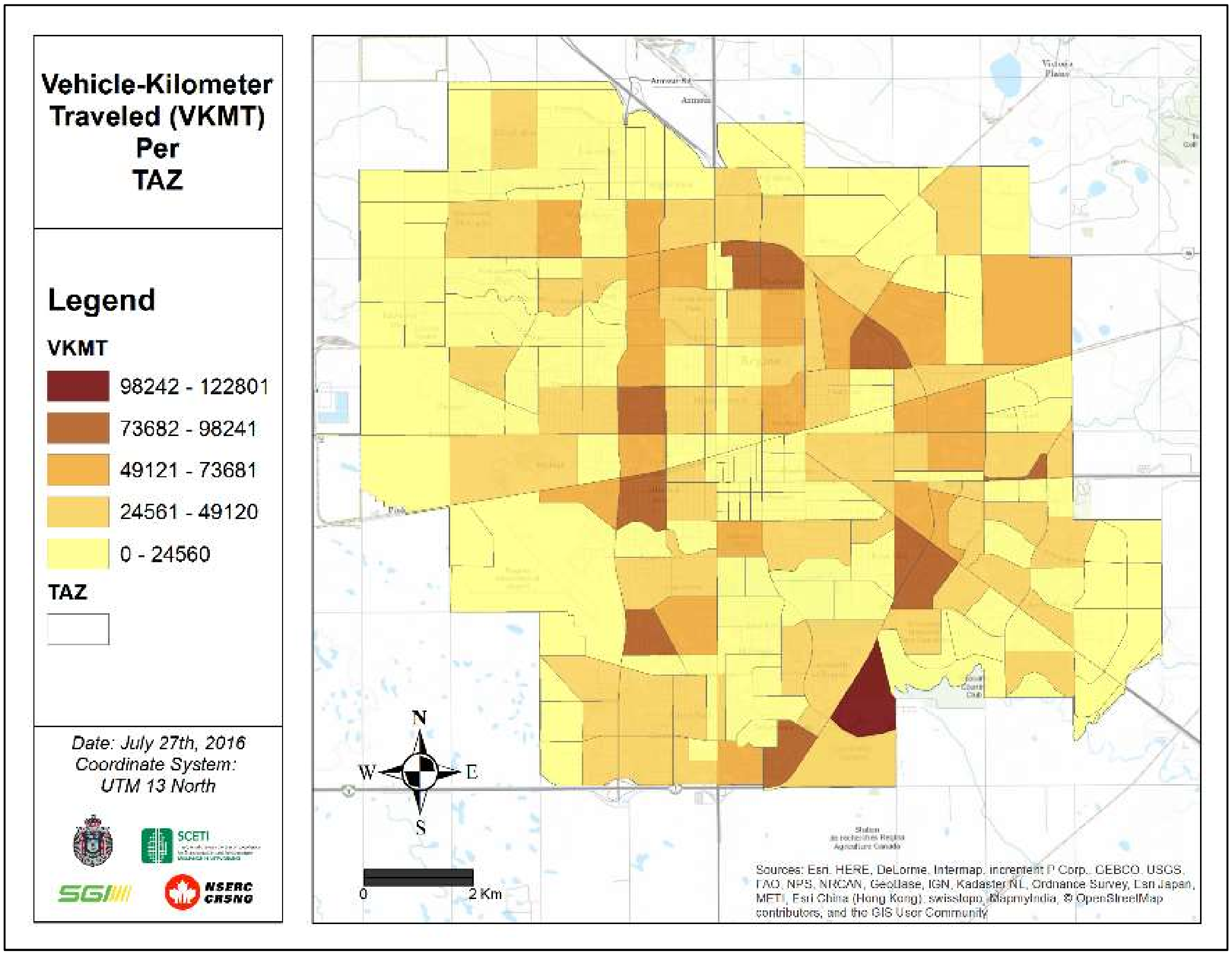
Number of Five-Leg Intersections Per TAZ



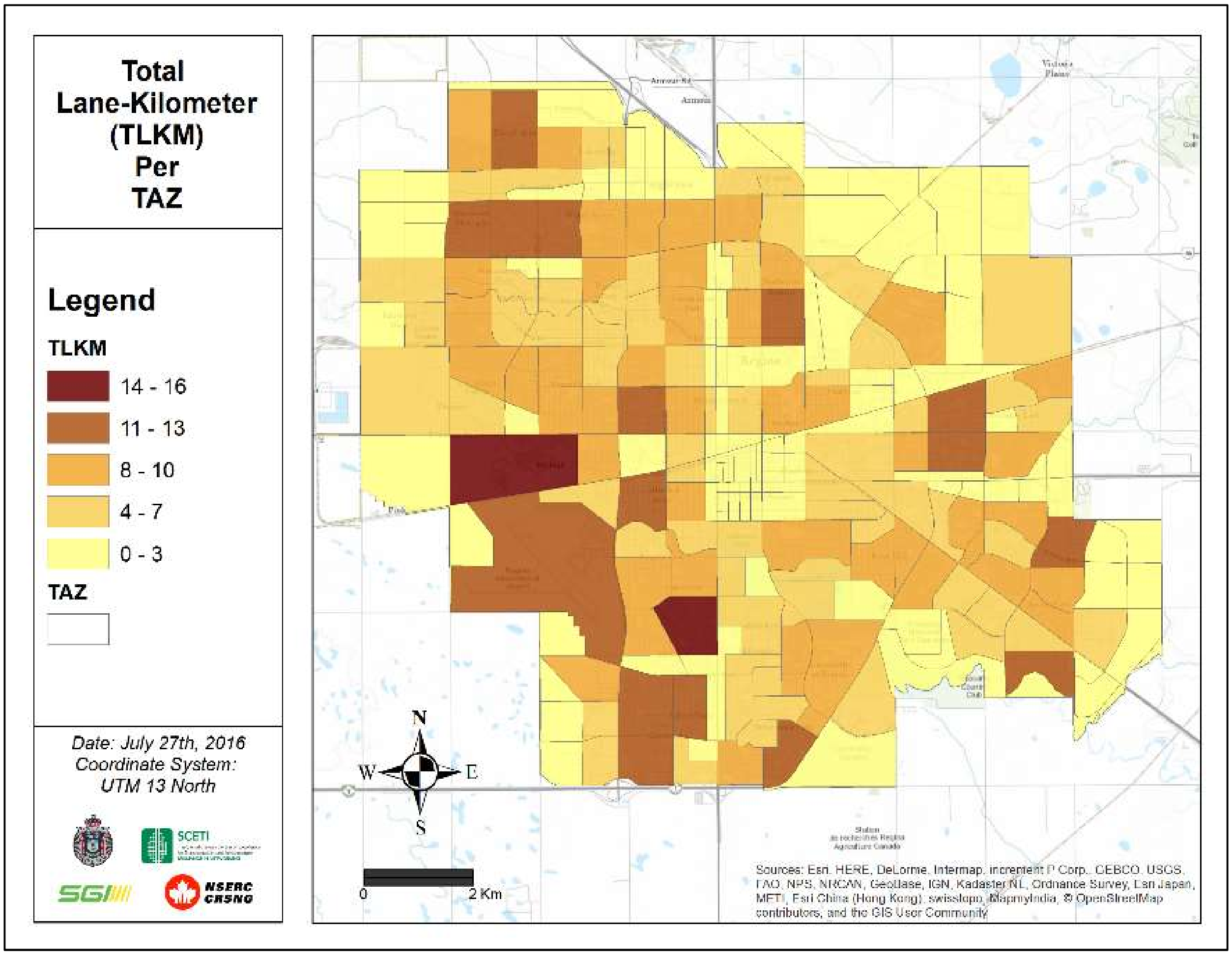
Total Number of Intersections Per TAZ



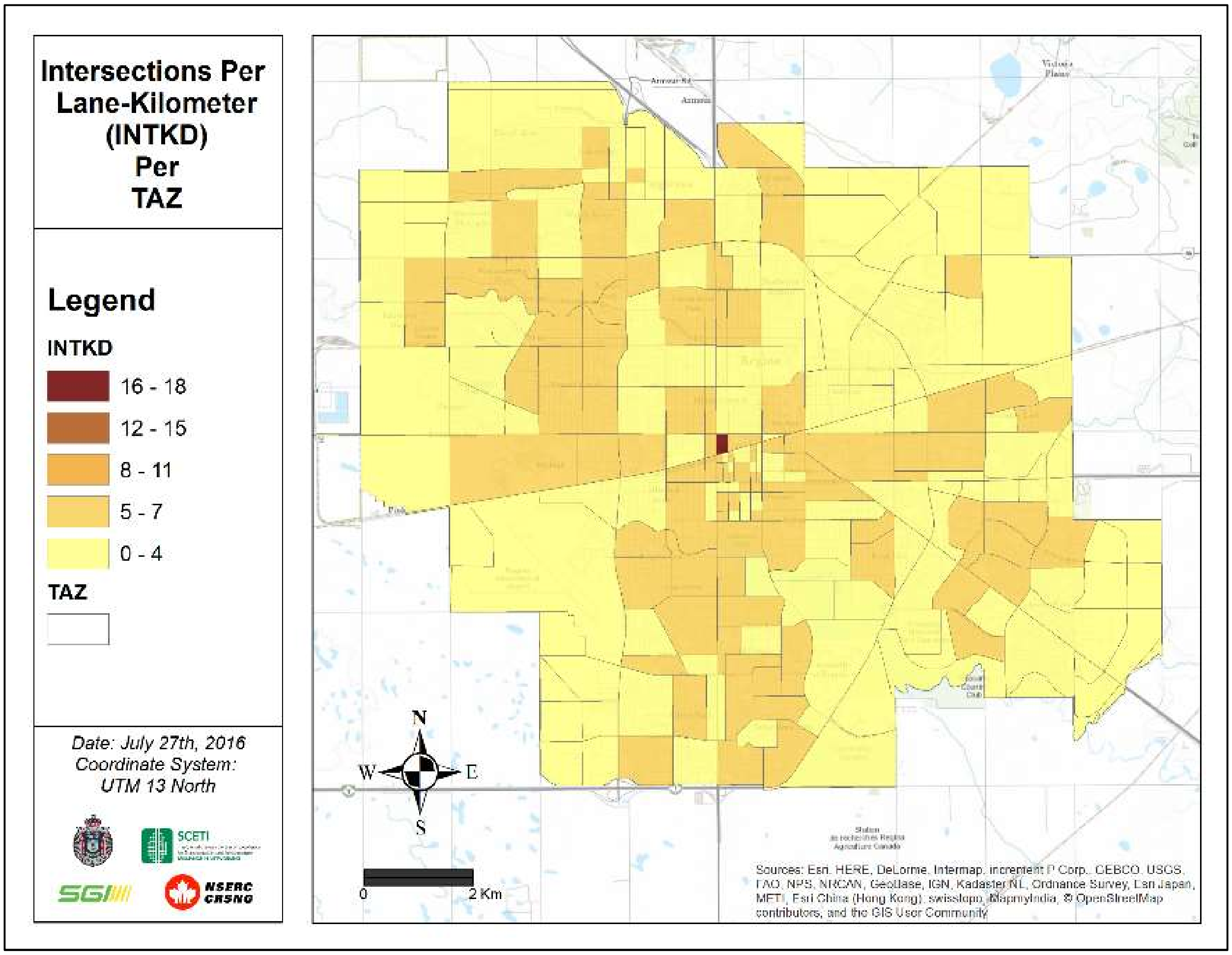
Vehicle-Kilometer Traveled Per TAZ



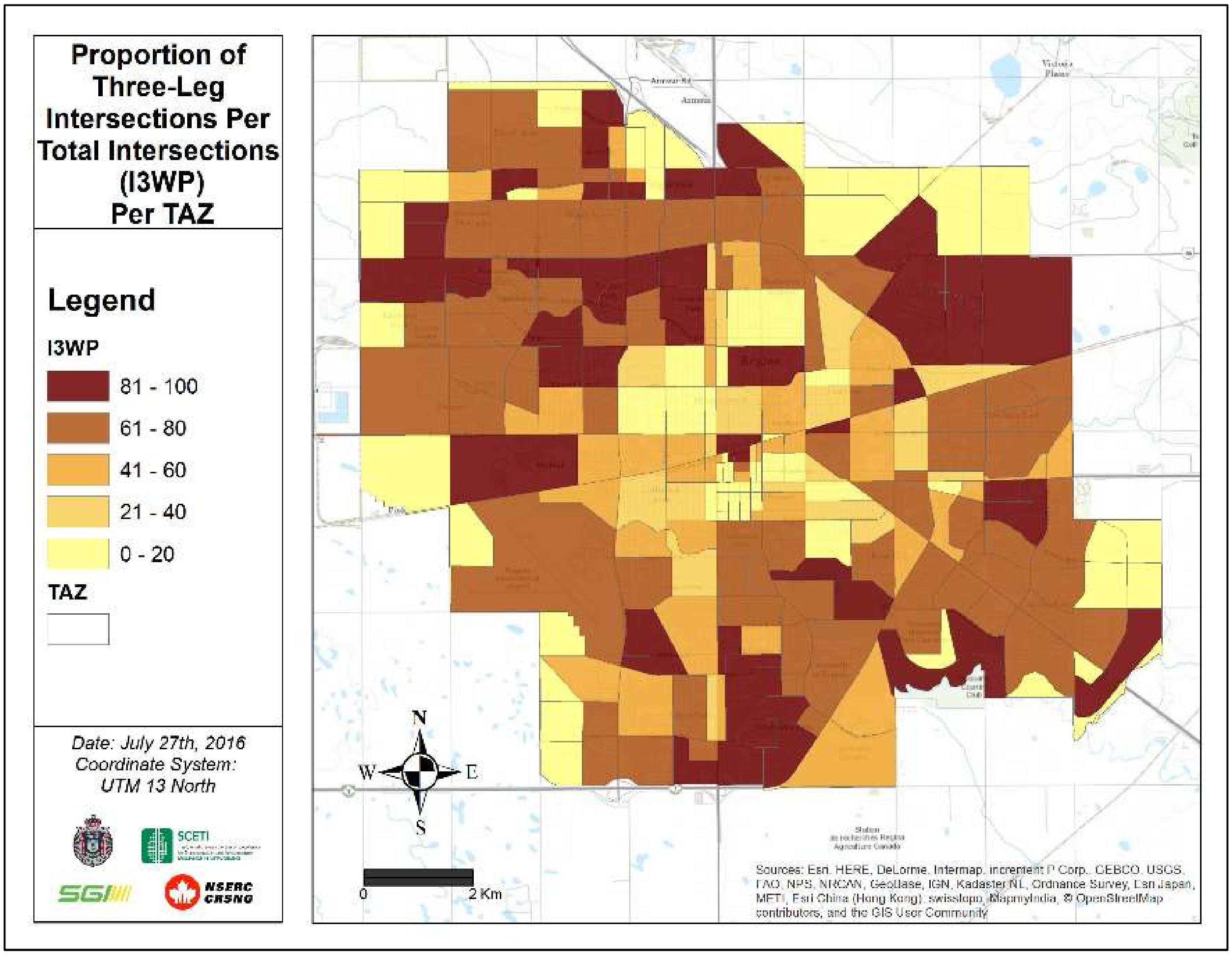
Total Lane-Kilometer Traveled (TLKM) Per TAZ (km)



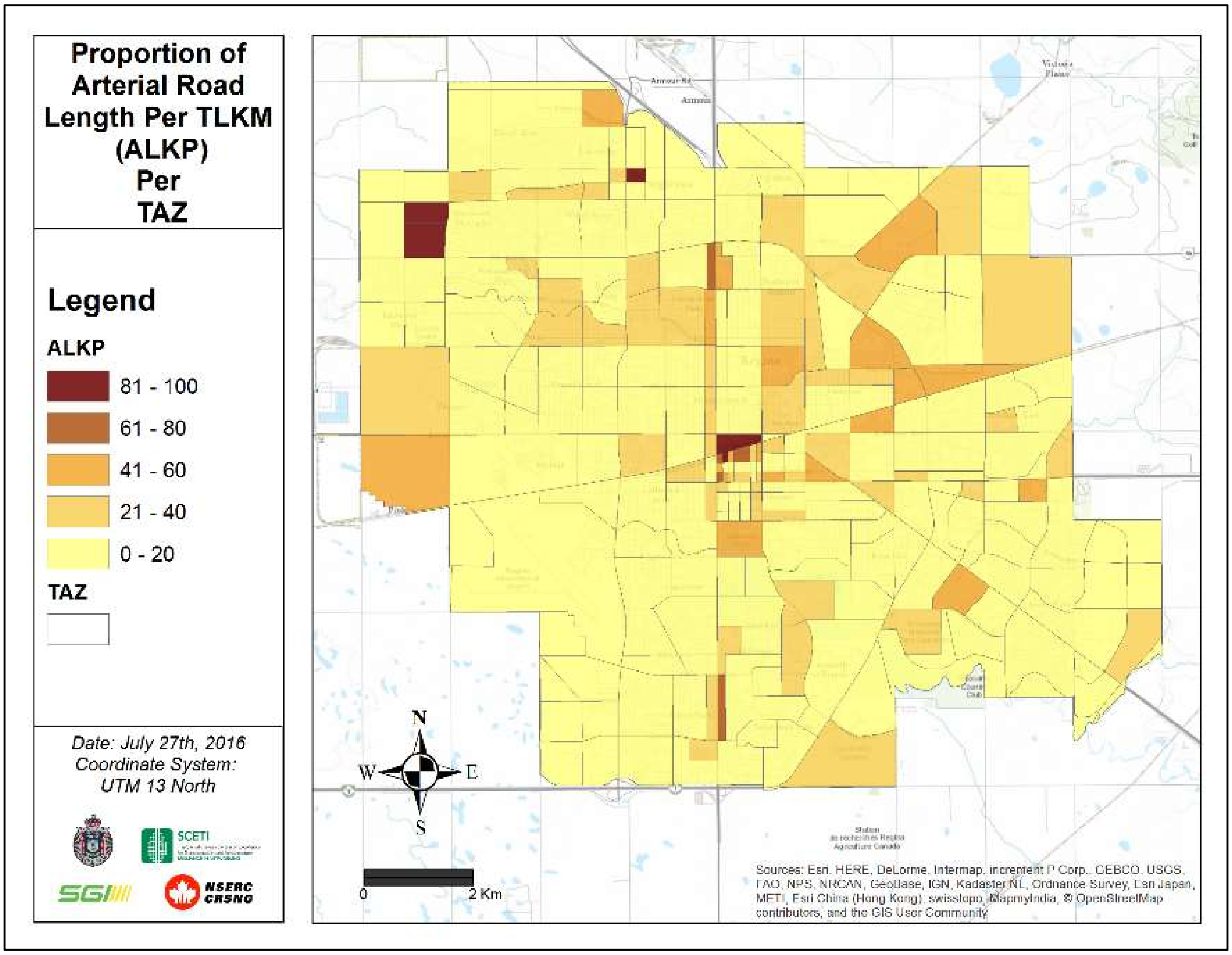
Number of Intersections per Total Lane-Kilometer Traveled (INTKD) Per TAZ



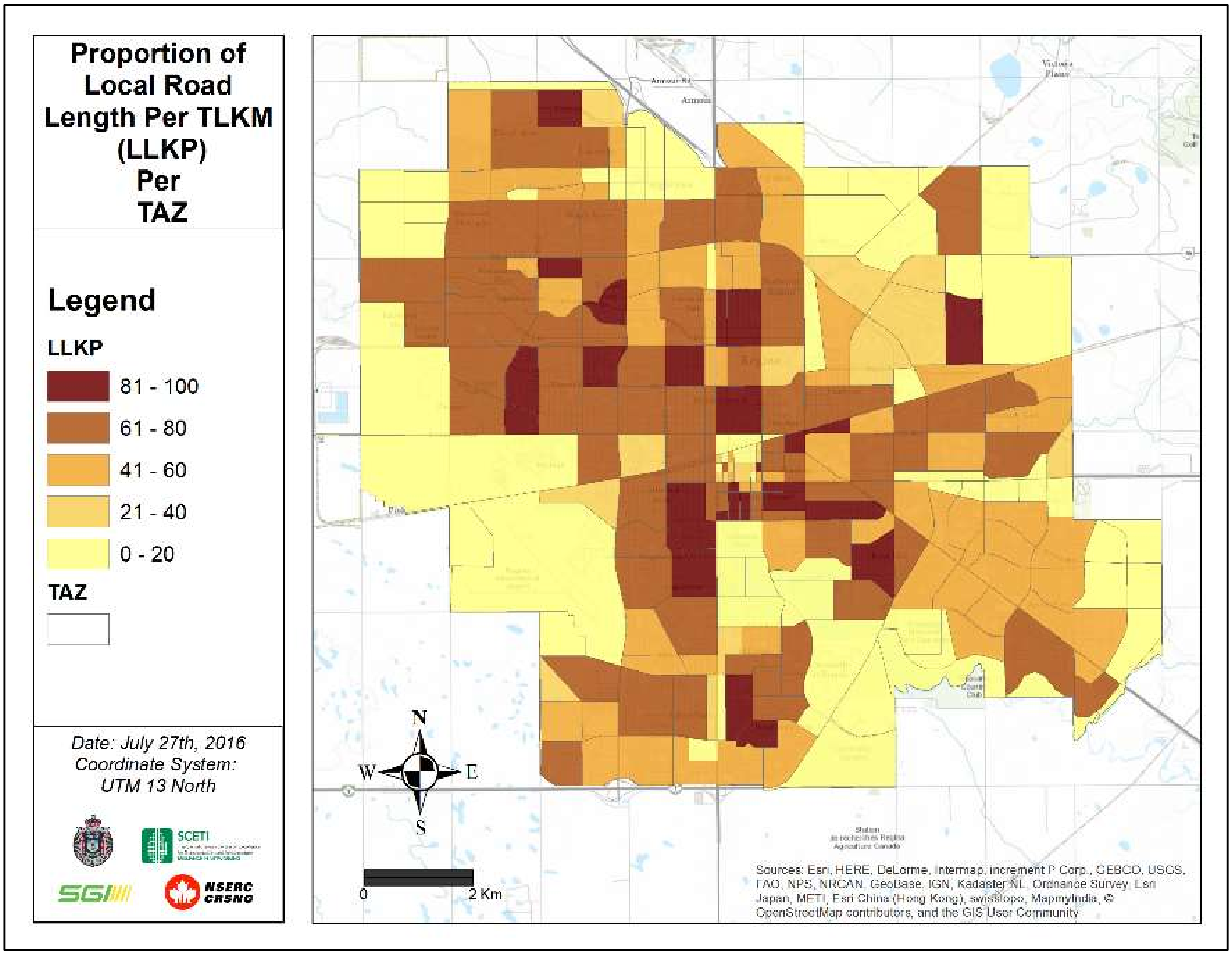
Proportion of Total Number of Intersections that are three-way (I3WP) (%)



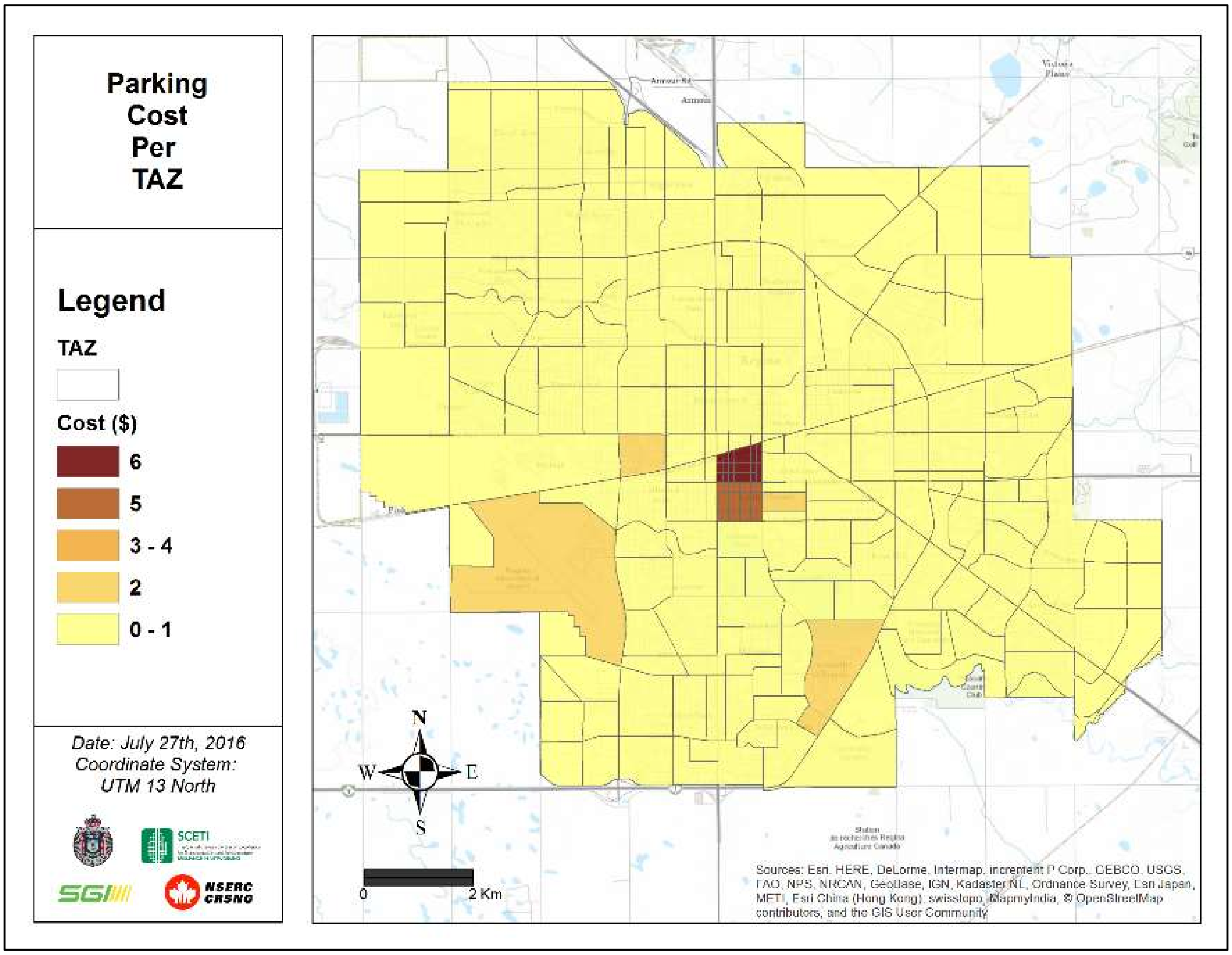
Proportion of Total Road Segment Length that are Arterial Roads (ALKP) (%)



Proportion of Total Road Segment Length that are Local Roads (LLKP) (%)



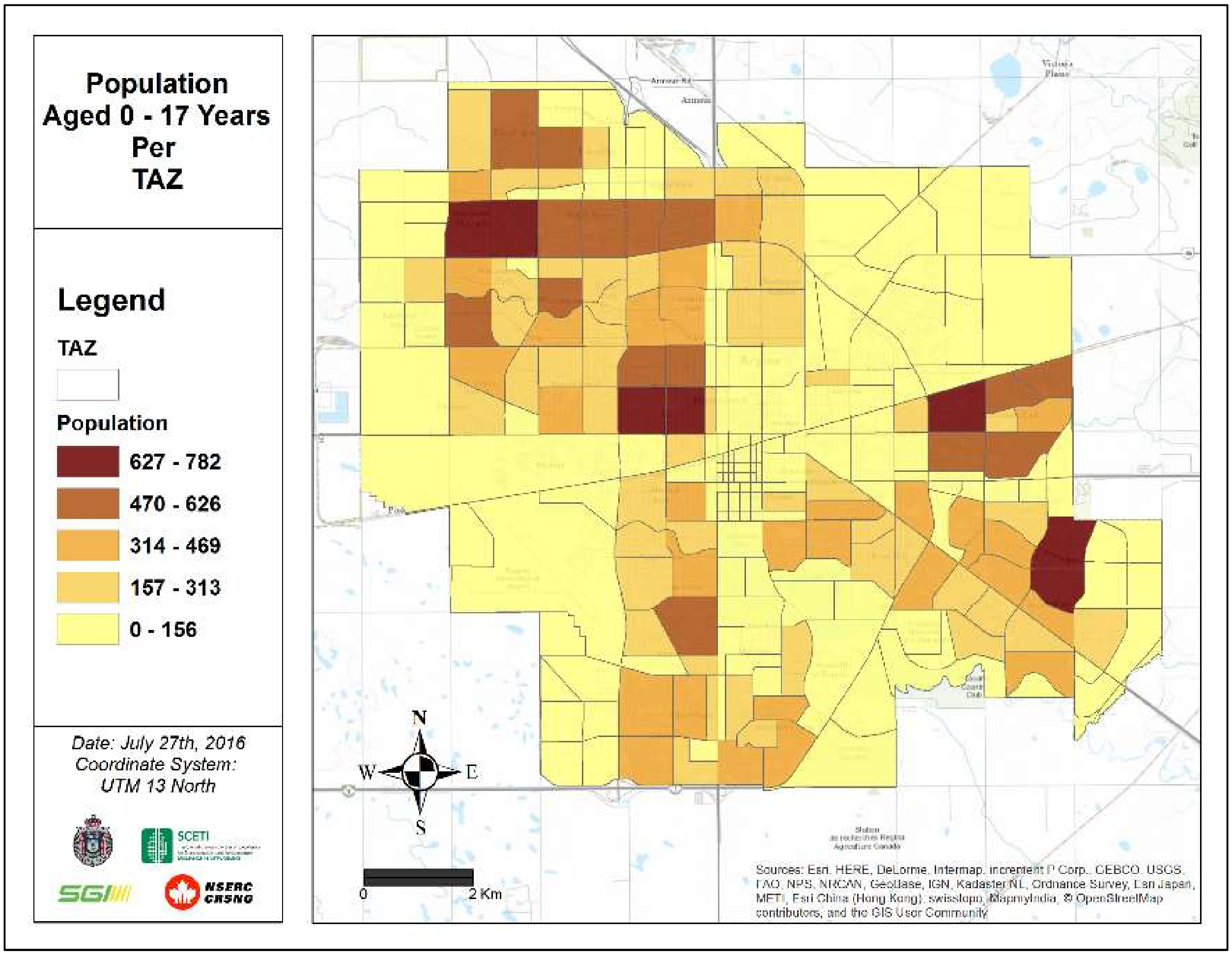
Parking Cost Per TAZ ($/hour)



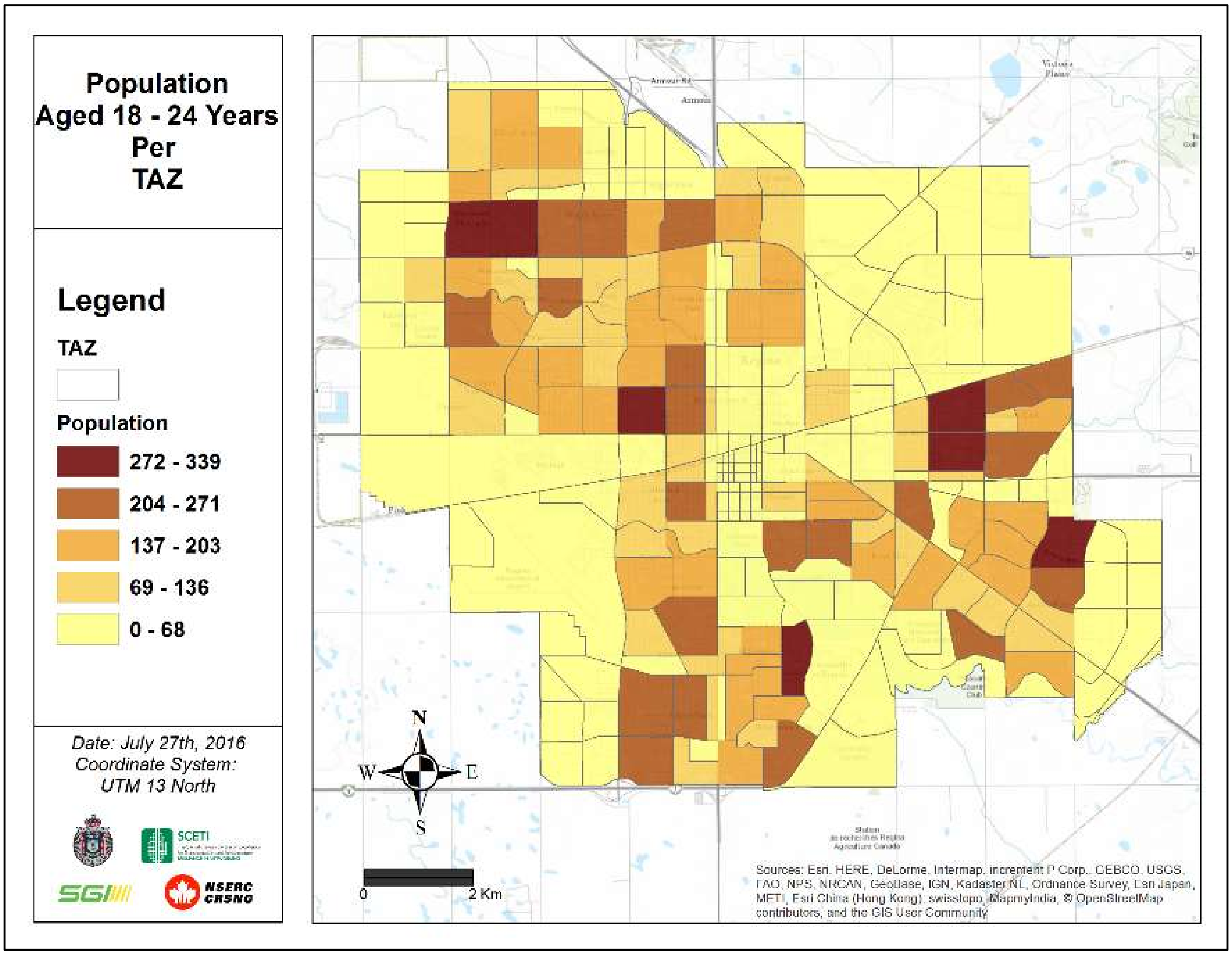
C5. Demographics

Variable	Observations	Total	Minimum	Maximum	Mean	Standard Deviation
Population aged 1 to 17	263	42431	0	782	161	179.45
Population aged 18 to 24	263	21145	0	339	80	84.12
Population aged 25 to 44	263	56461	0	884	215	220.58
Population aged 45 to 64	263	53726	0	911	204	214.47
Population aged 65 and above	263	27443	0	763	104	125.84
Total Population	263	201218	0	3011	765	785.32
Number of graduate students	263	29367	0	1720	112	261.35

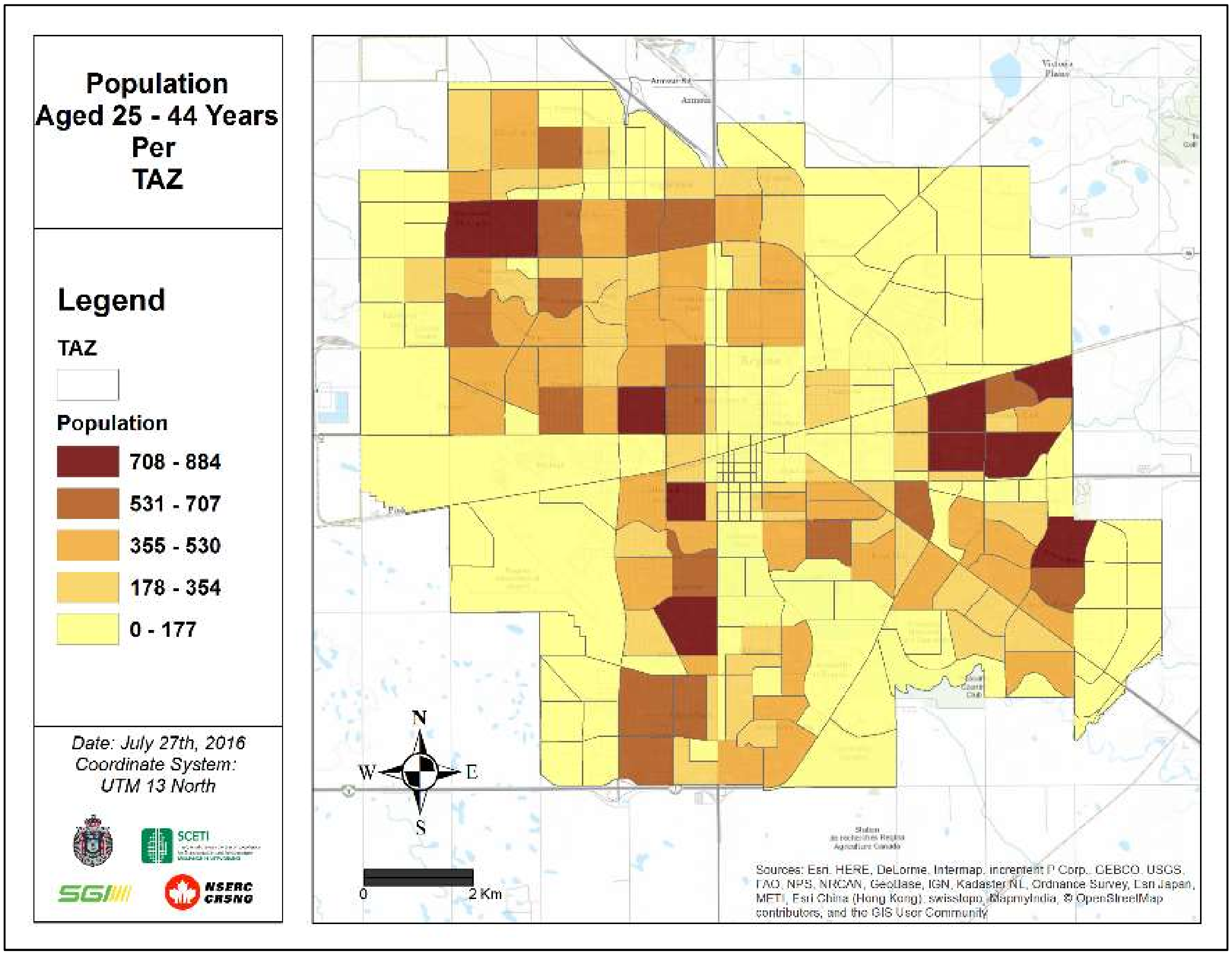
Number of Residents Aged 0 to 17 years Per TAZ



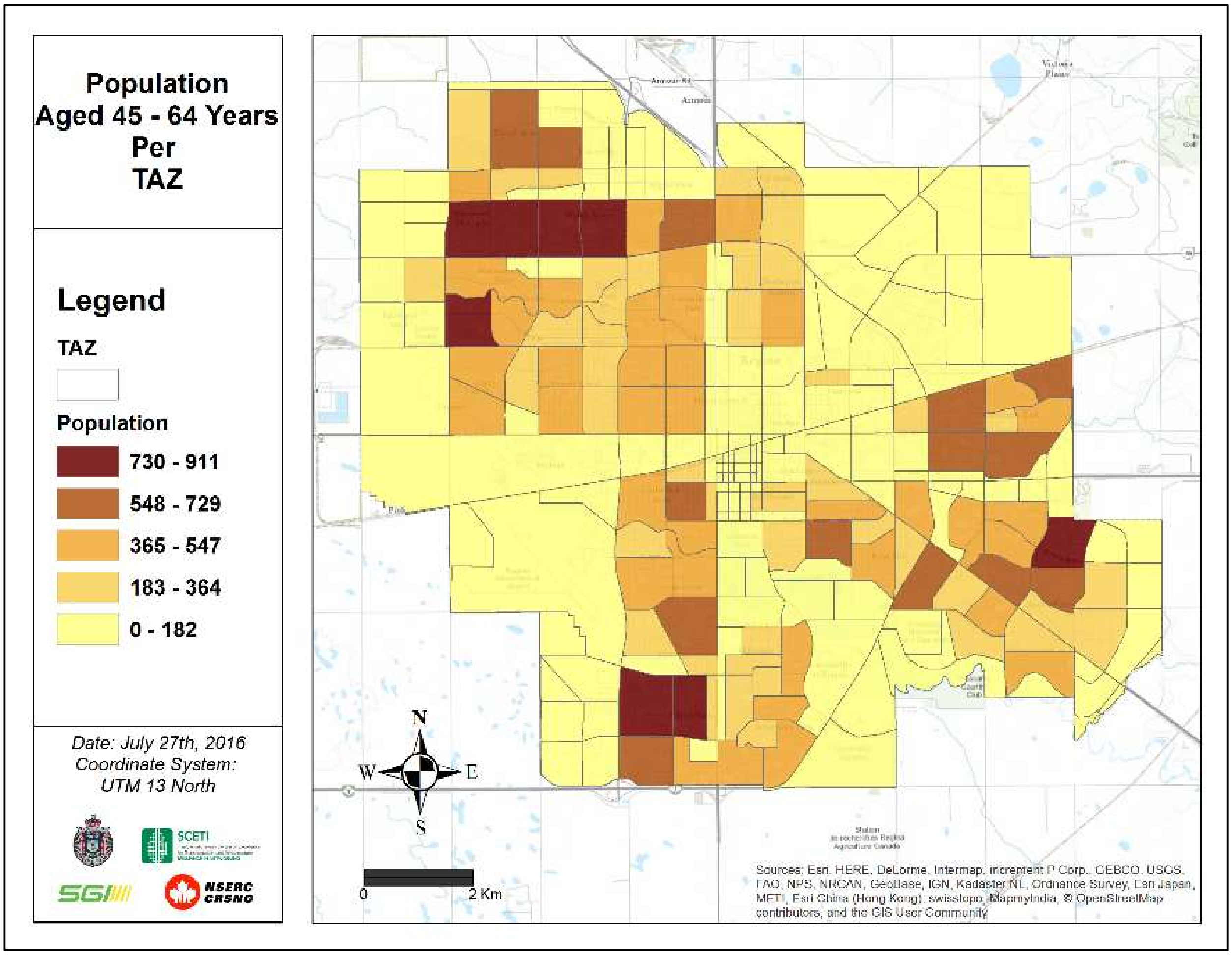
Number of Residents Aged 18 to 24 years Per TAZ



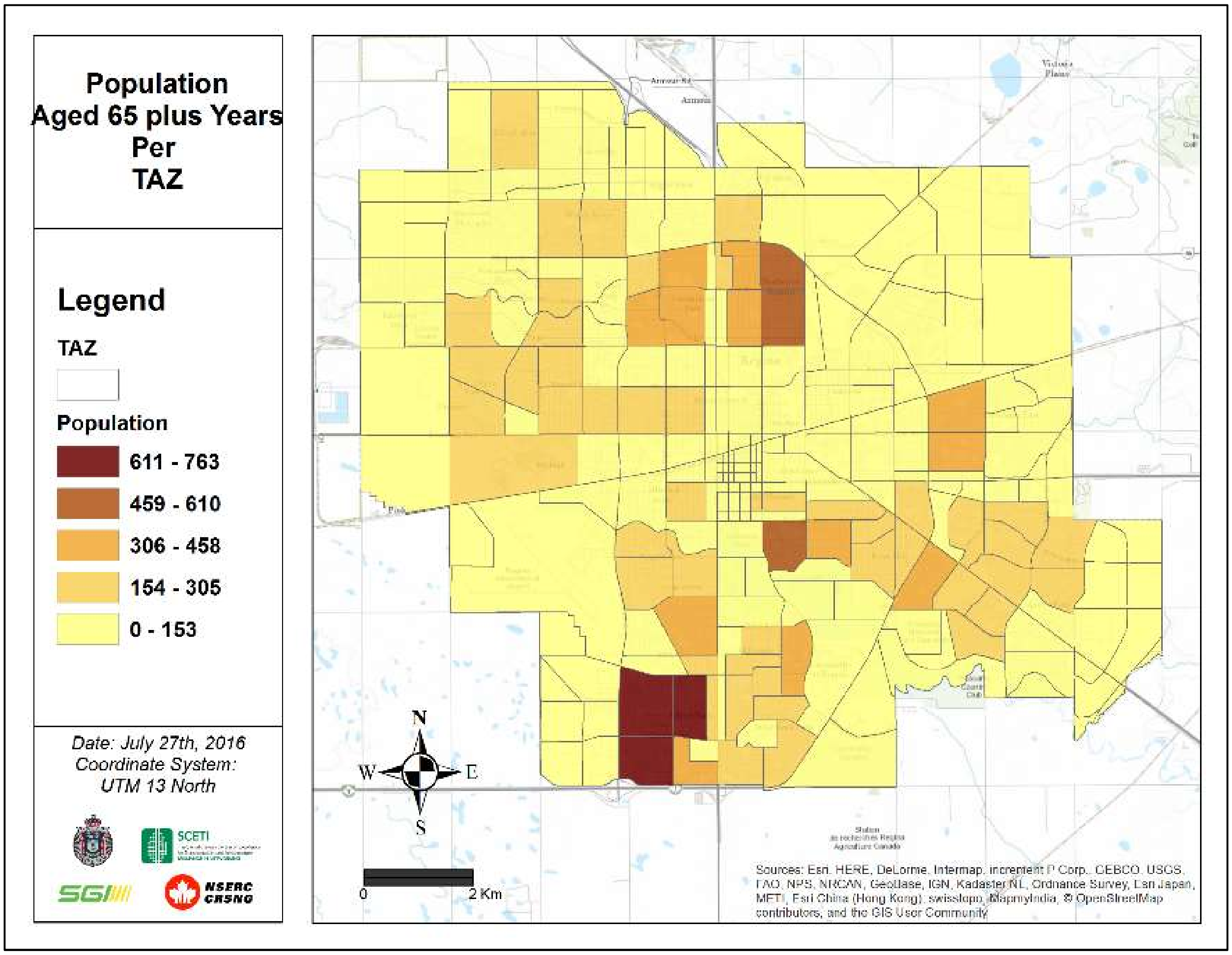
Number of Residents Aged 25 to 44 years Per TAZ



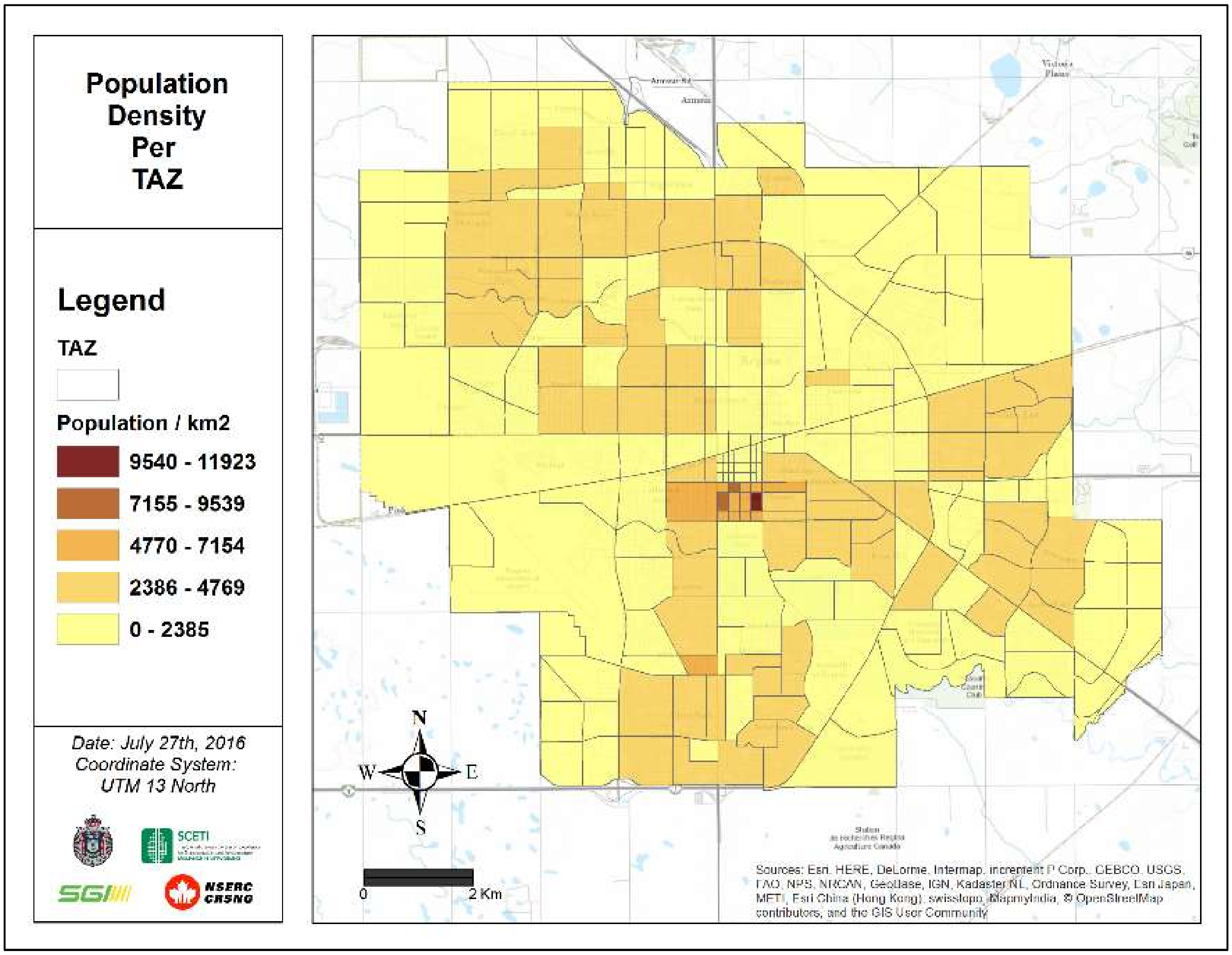
Number of Residents Aged 45 to 64 years Per TAZ



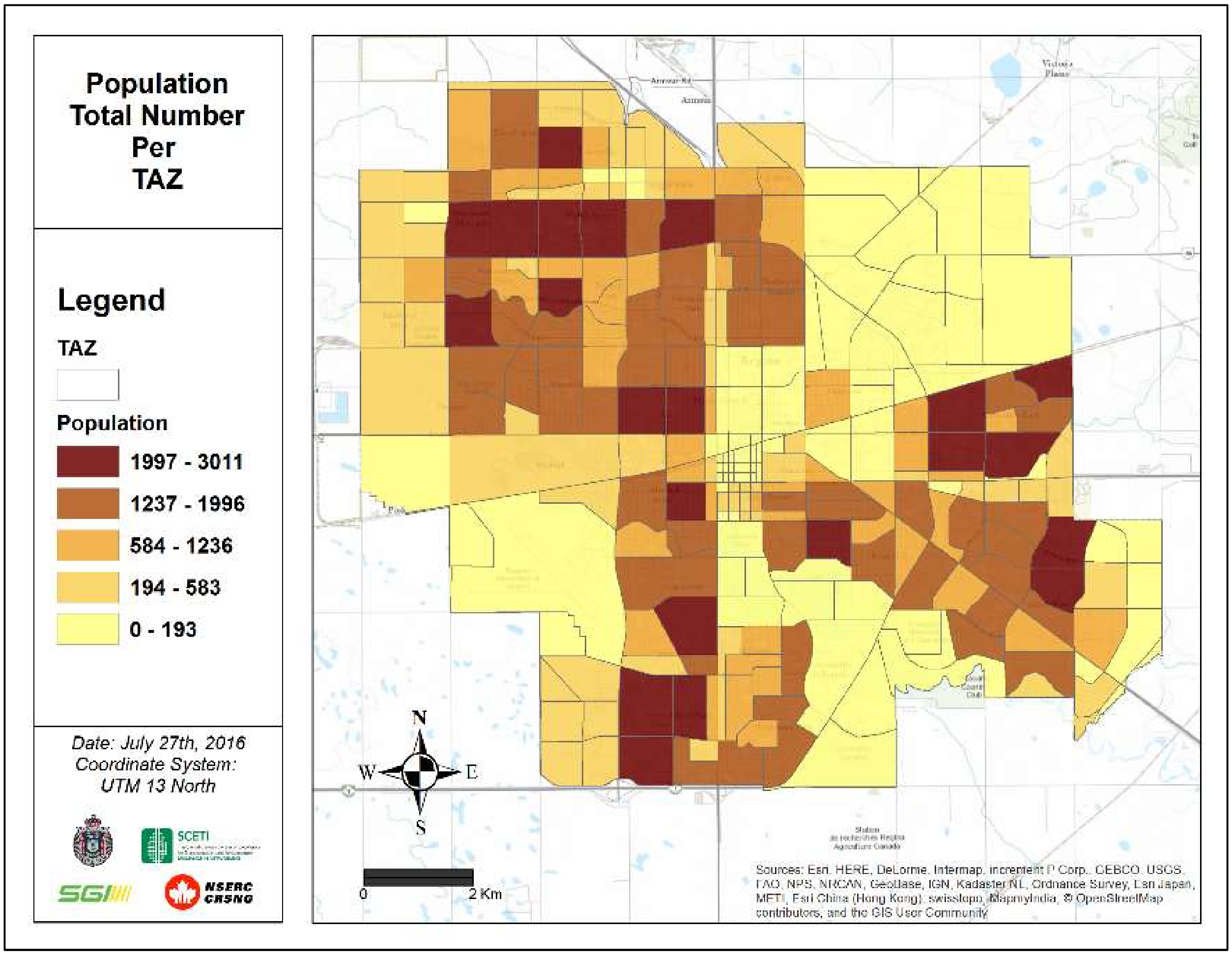
Number of Residents Aged 65 years and above Per TAZ



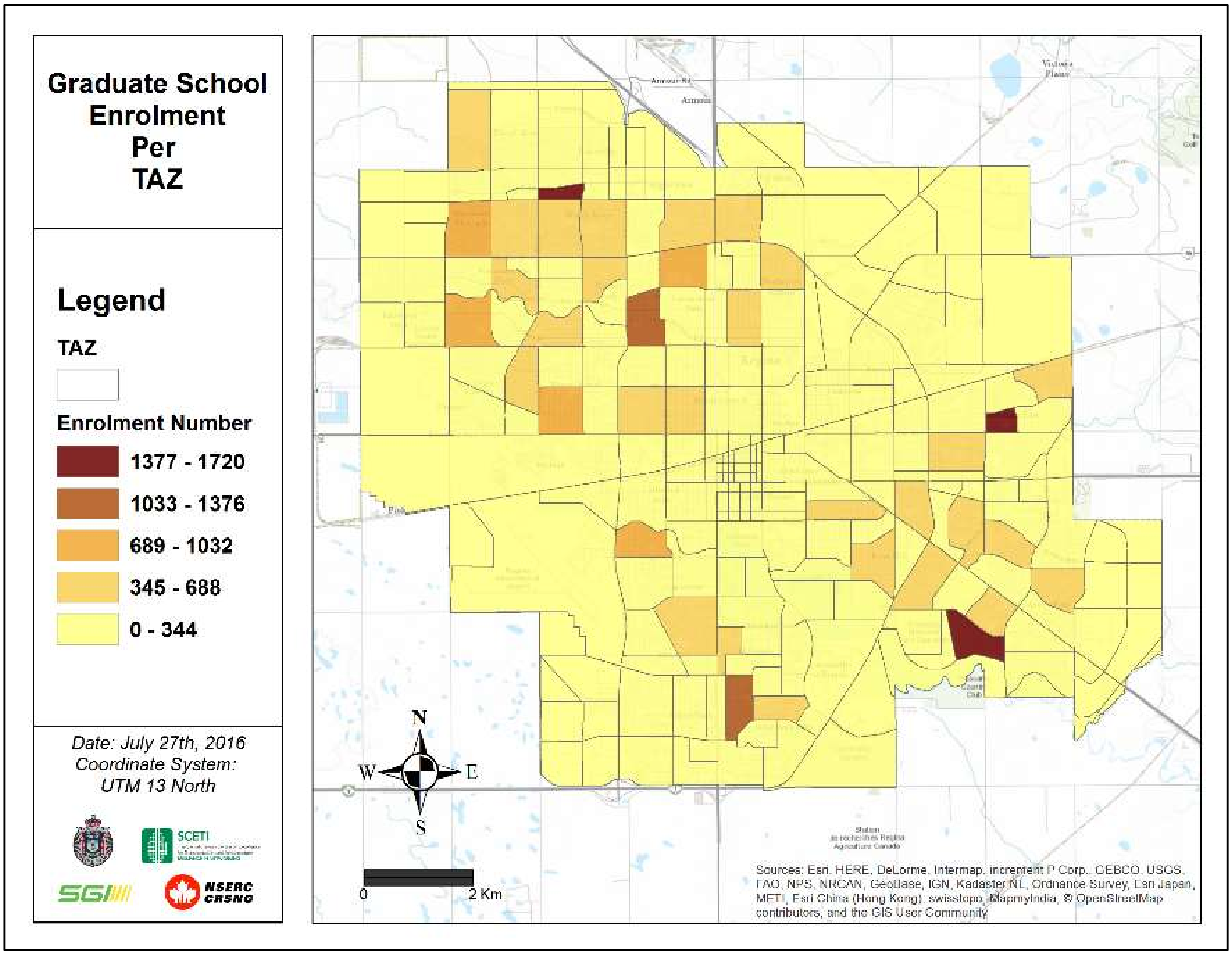
Population Density Per TAZ (Persons/km^2)



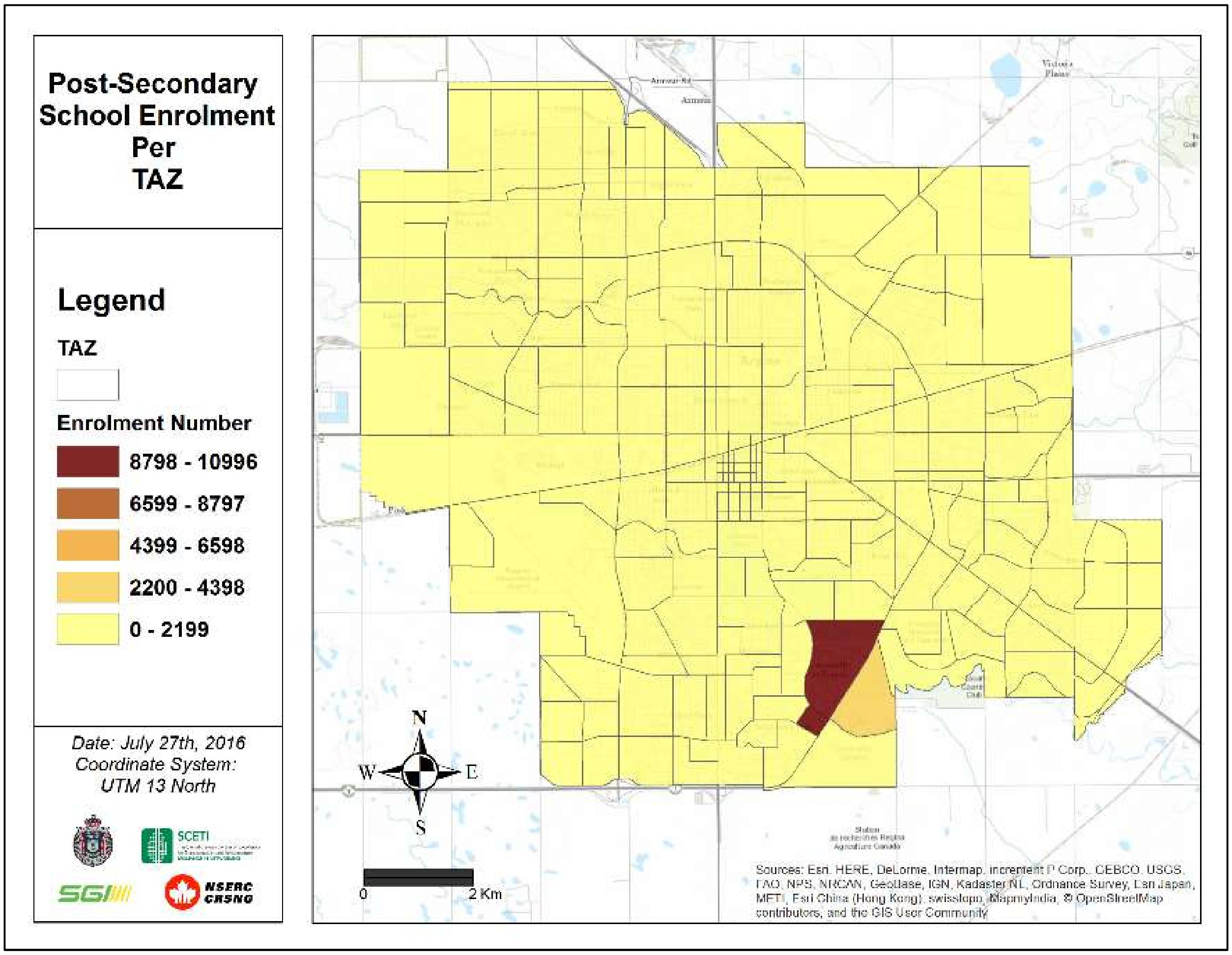
Total Number of Residents Per TAZ



Total Number of Residents Enrolled in Graduate School Per TAZ



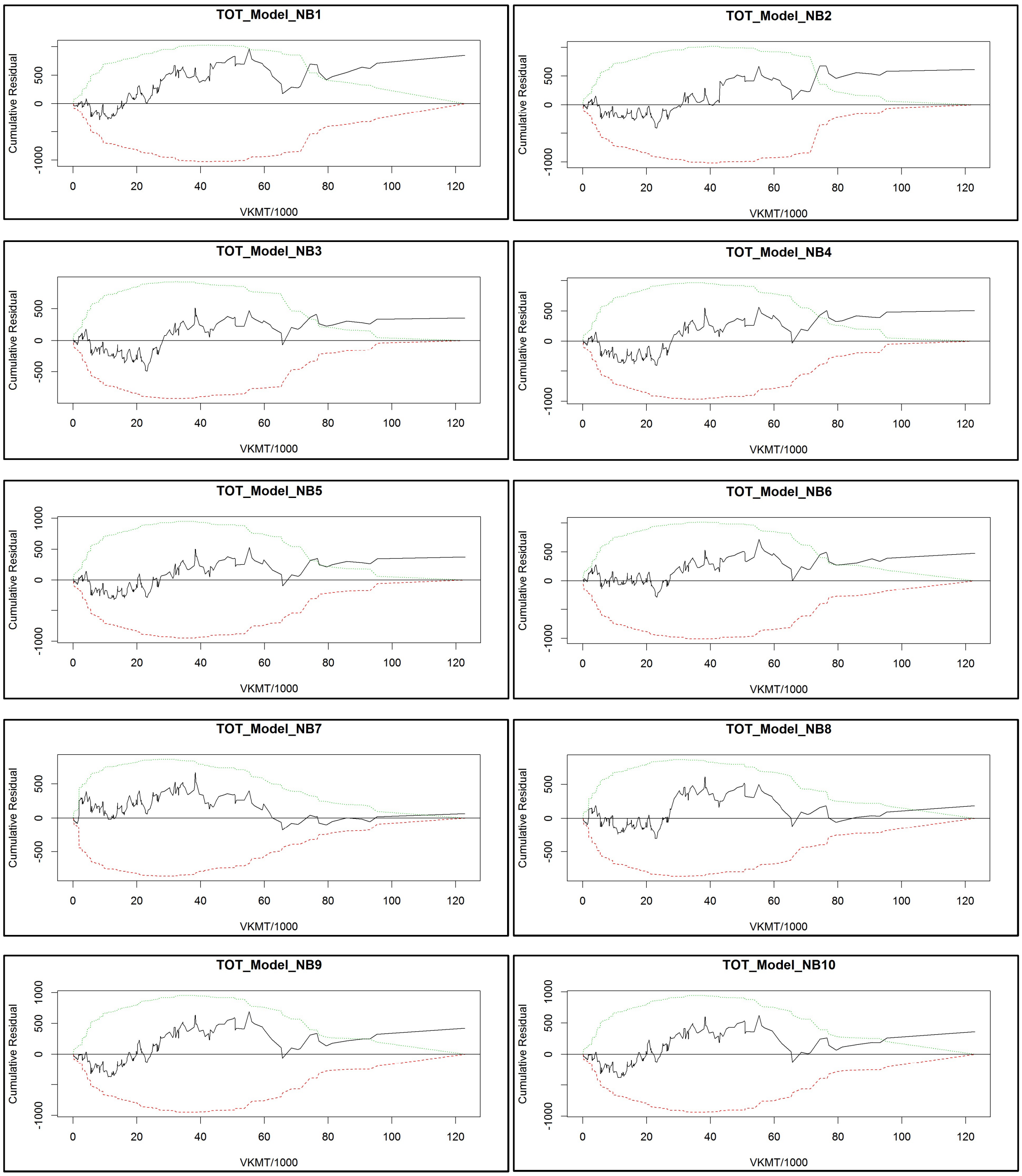
Total Number of Residents Enrolled in a Post Secondary School Per TAZ



APPENDIX D: Cumulative Residual Plots for Candidate Models

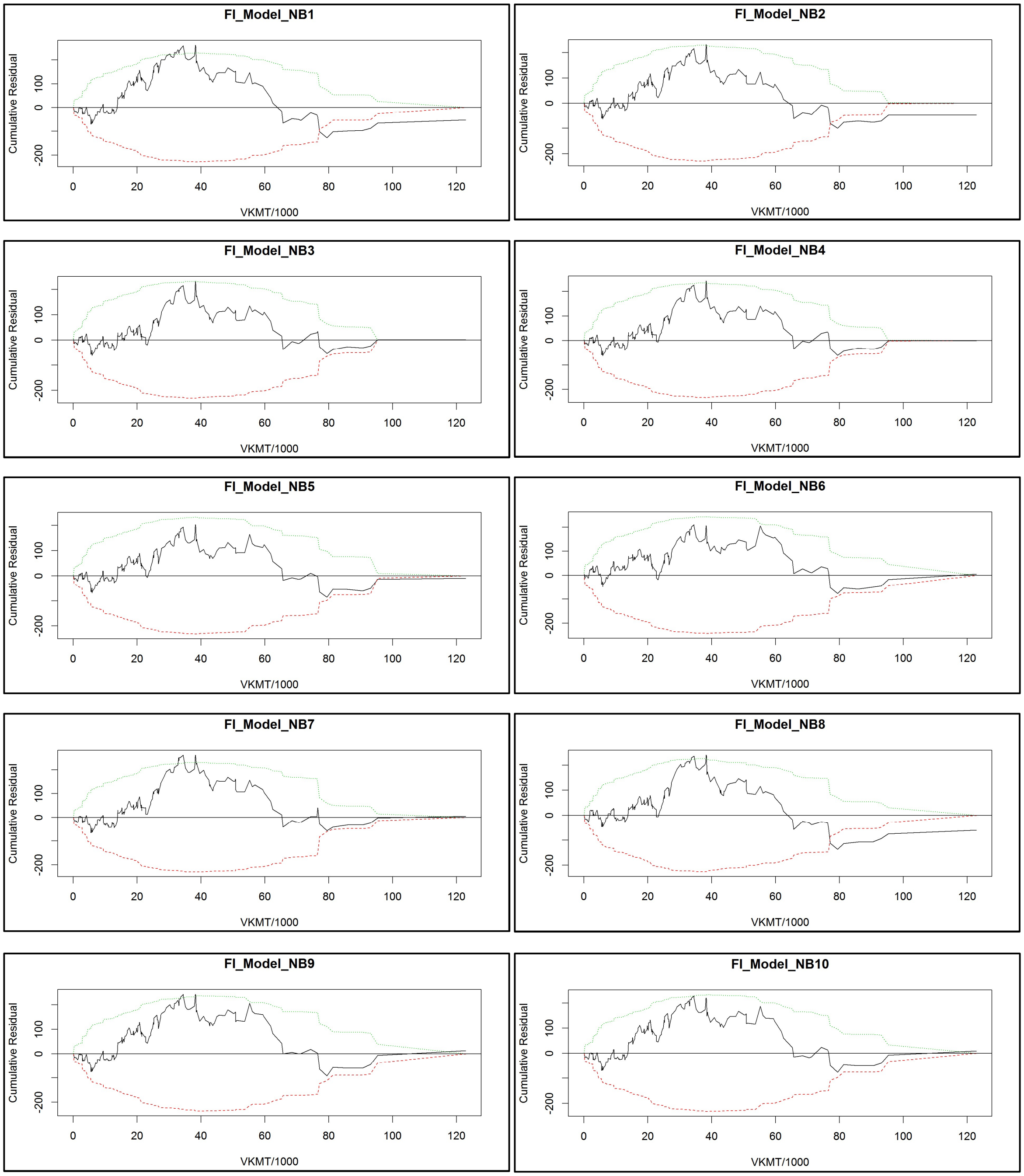
D1: Total Collisions

Cumulative Residual Plots for Total Collisions Top 10 Models



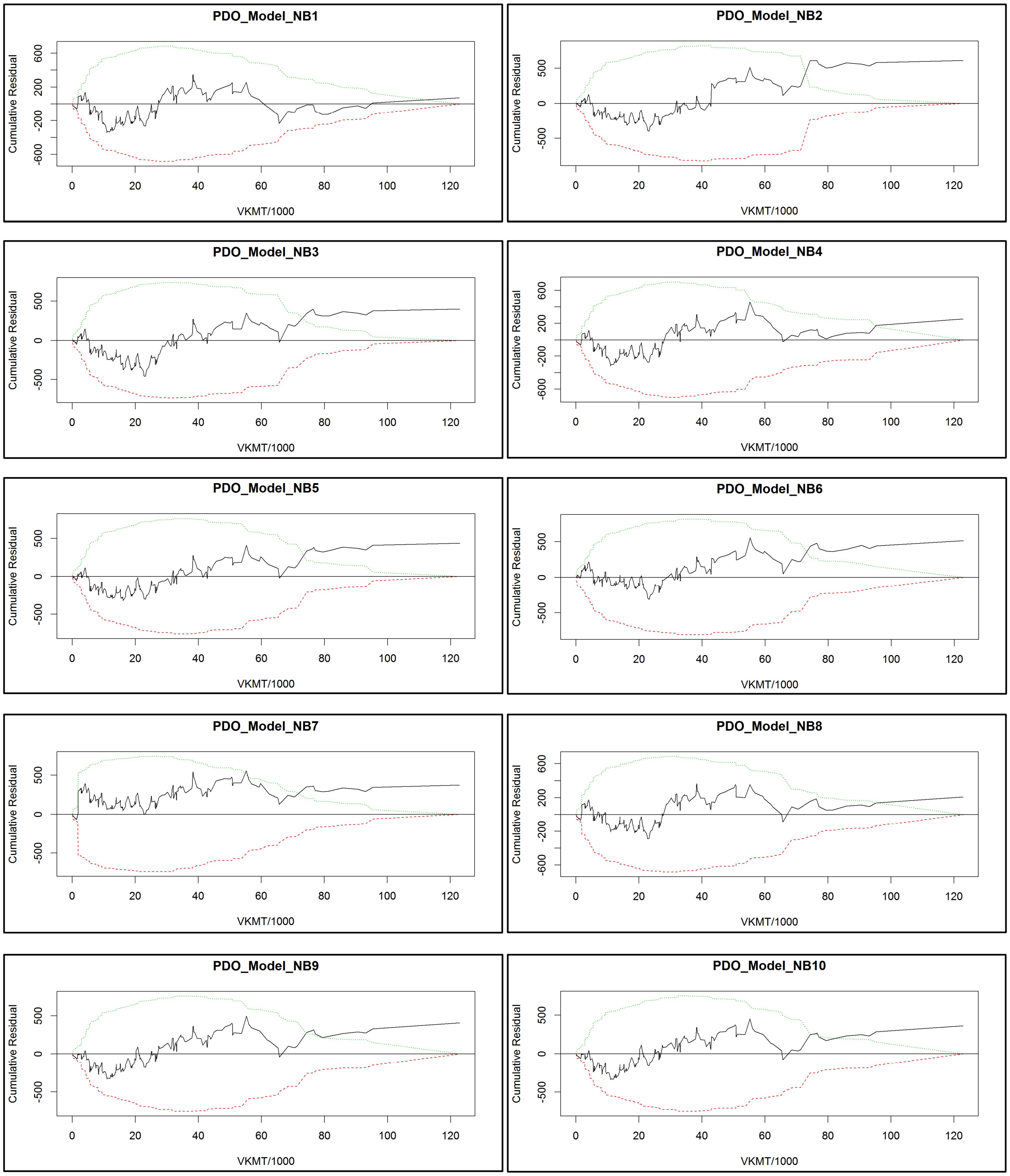
D2: Fatal-Injury Collisions

Cumulative Residual Plots for Fatal-Injury Collisions Top 10 Models



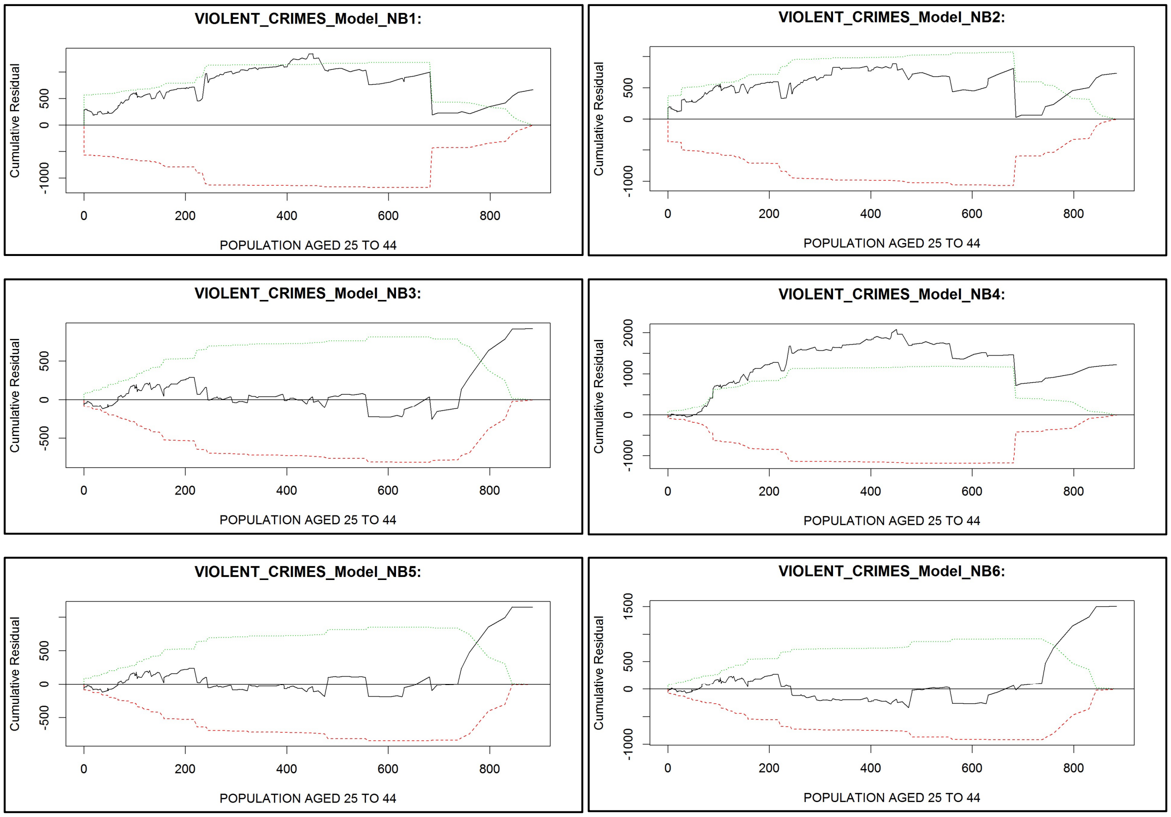
D3: Property Damage Only Collisions

Cumulative Residual Plots for Property Damage Only Collisions Top 10 Models



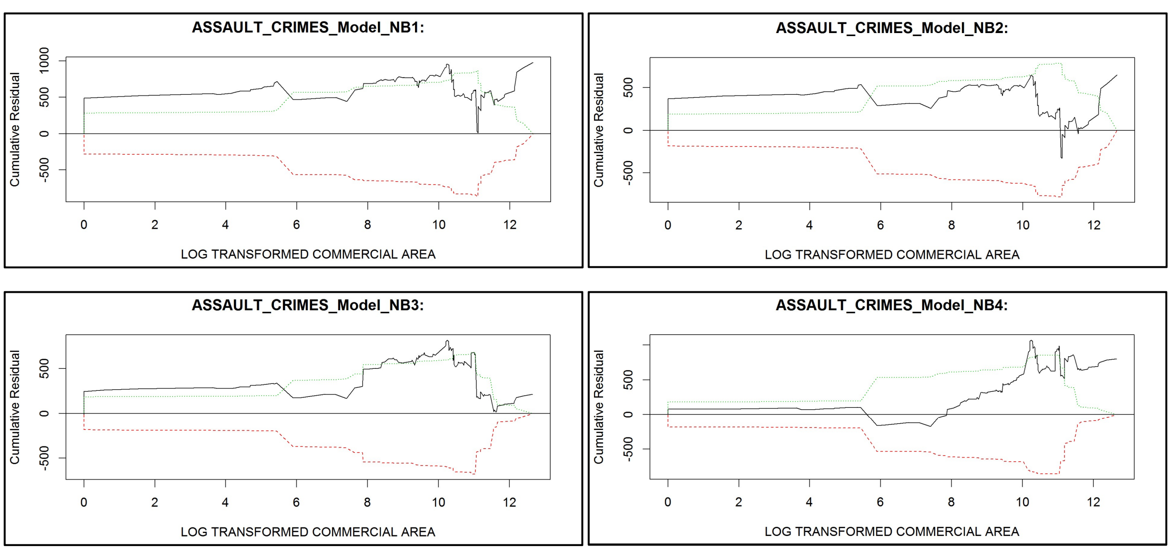
D4: Violent Crimes

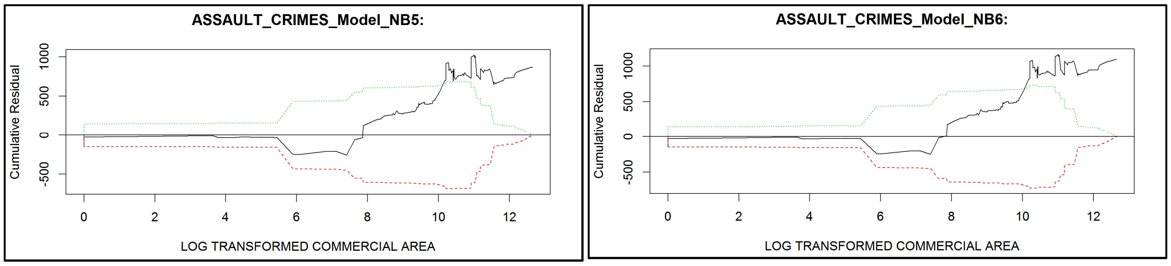
Cumulative Residual Plots for Violent Crimes Top 6 Models



D5: Assault Crimes

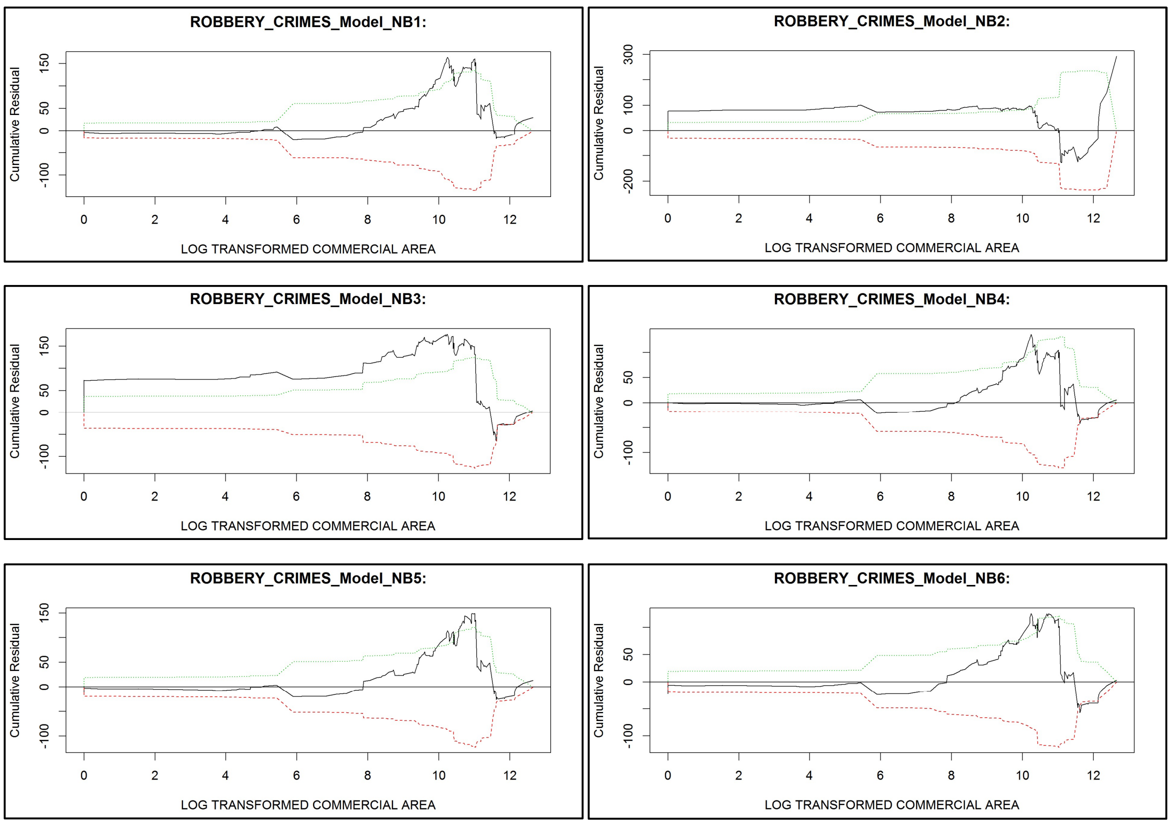
Cumulative Residual Plots for Assault Crimes Top 6 Models





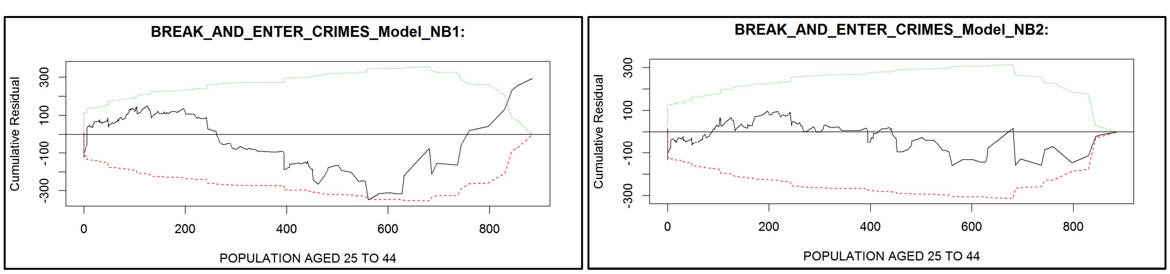
D6: Robbery Crimes

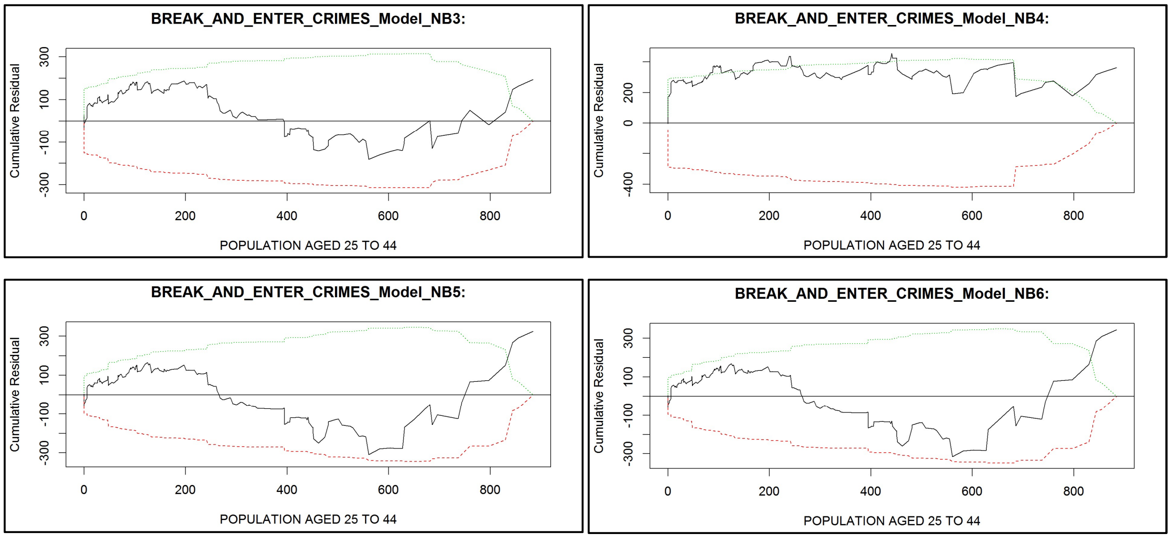
Cumulative Residual Plots for Robbery Crimes Top 6 Models



D7: Break and Enter Crimes

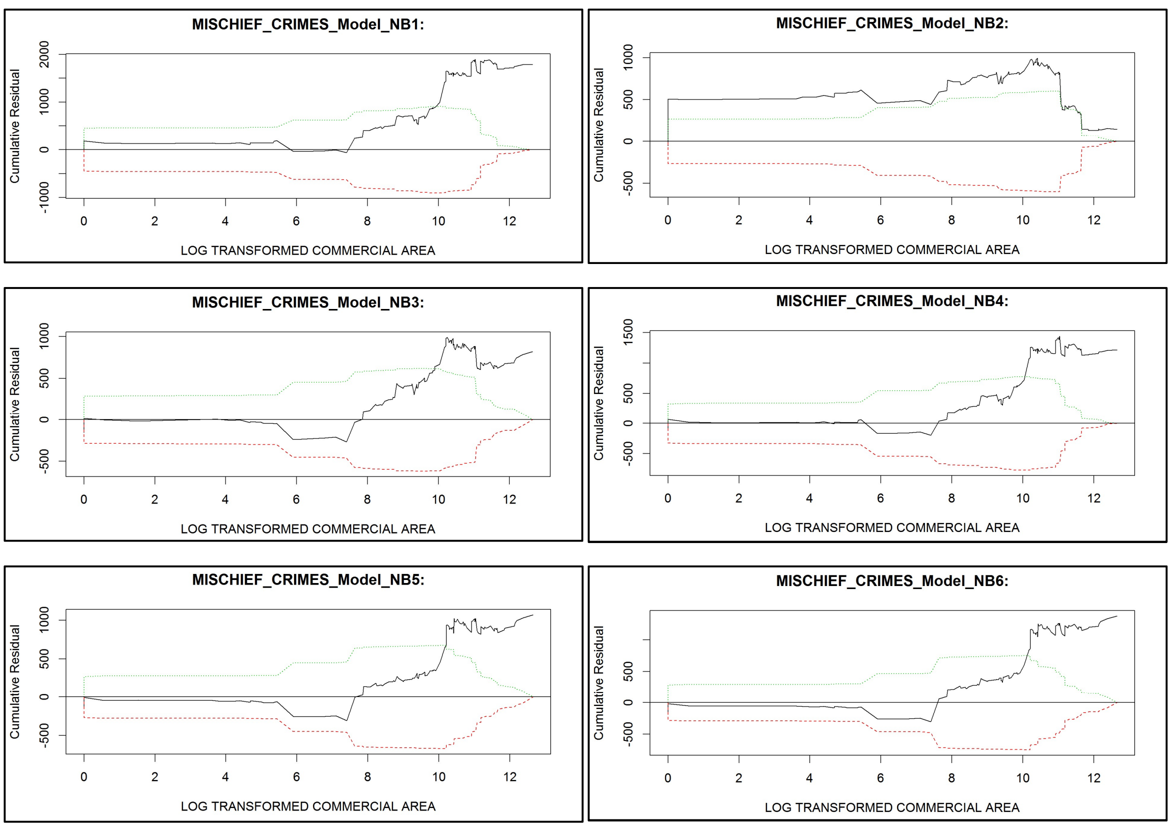
Cumulative Residual Plots for Break and Enter Crimes Top 6 Models





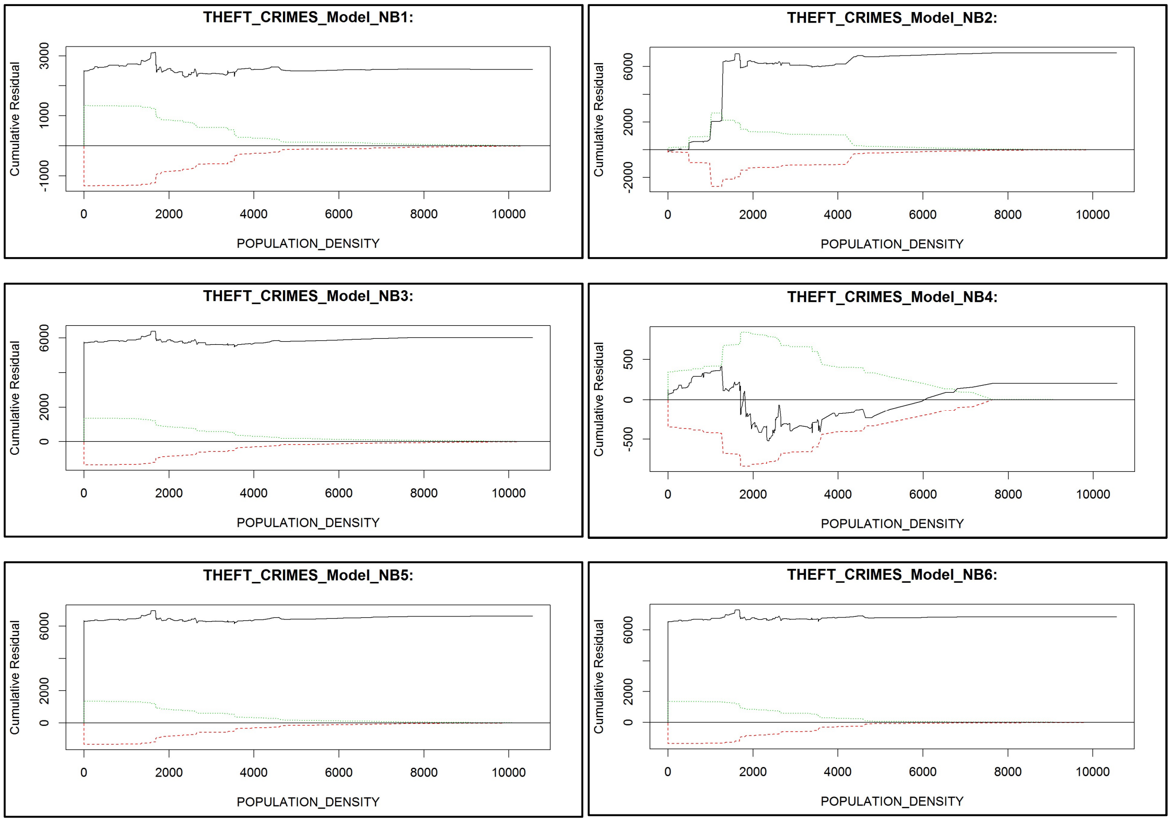
D8: Mischief Crimes

Cumulative Residual Plots for Mischief Crimes Top 6 Models



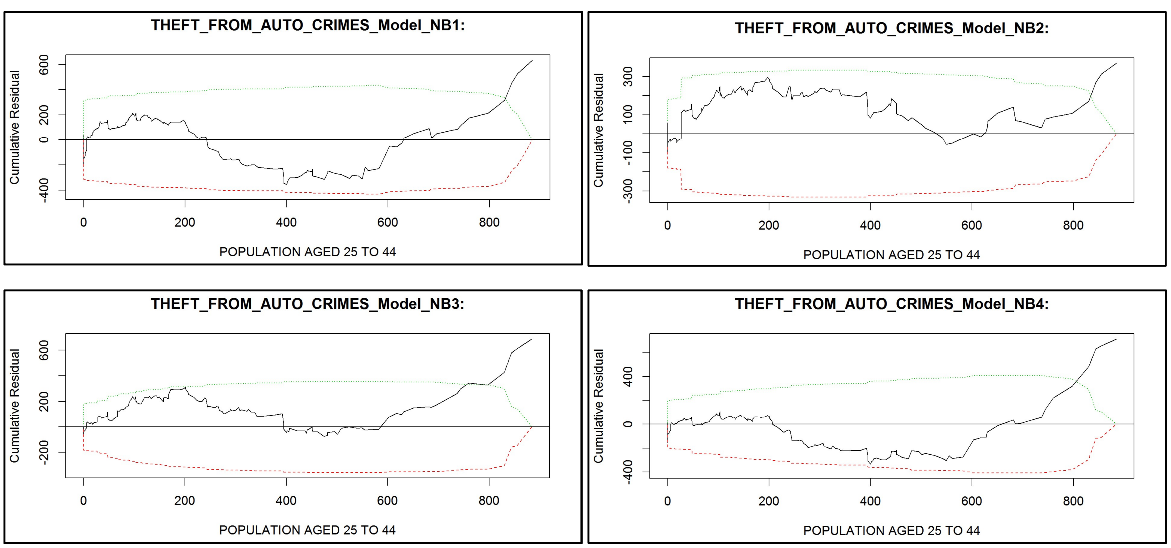
D9: Theft Crimes

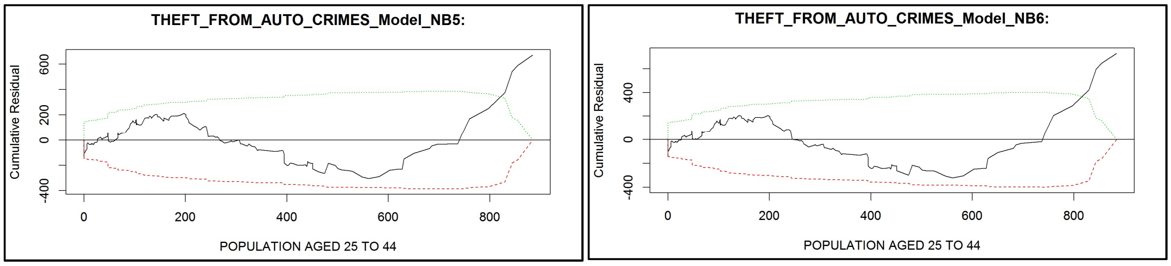
Cumulative Residual Plots for Theft Crimes Top 6 Models



D10: Theft from Auto

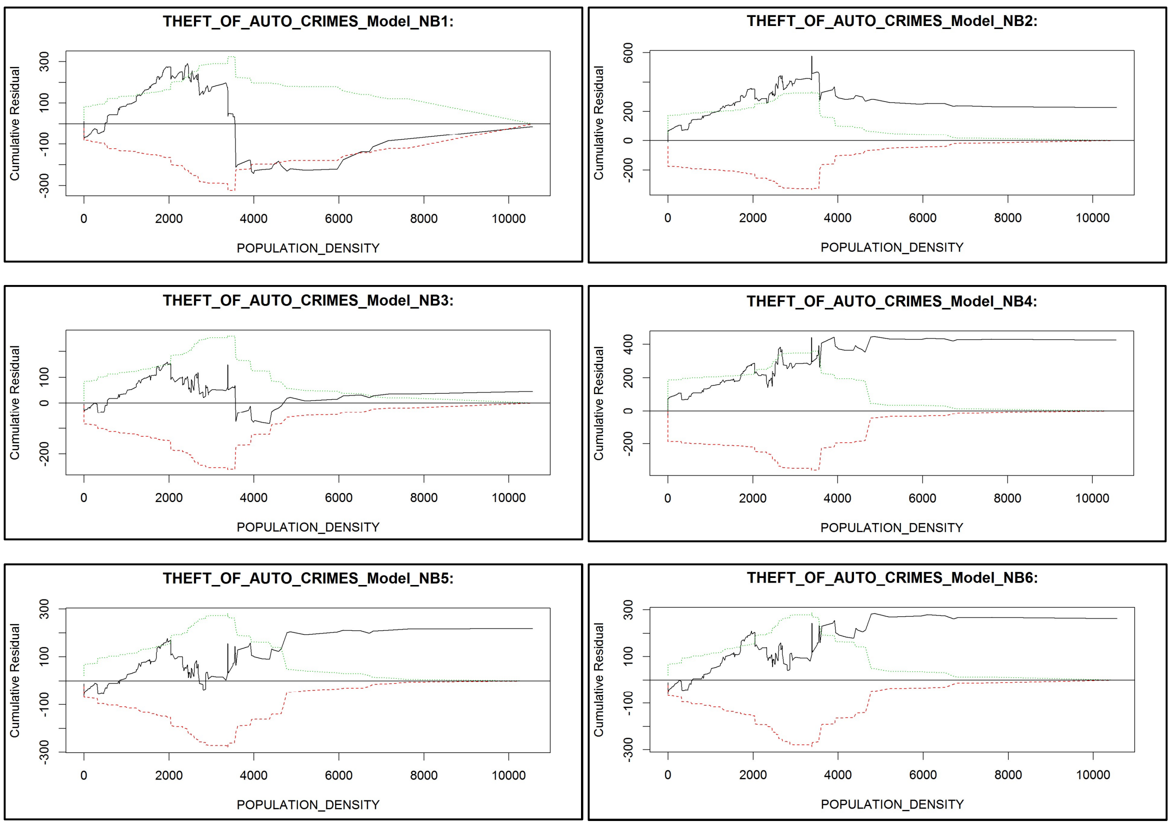
Cumulative Residual Plots for Theft from Auto Crimes Top 6 Models





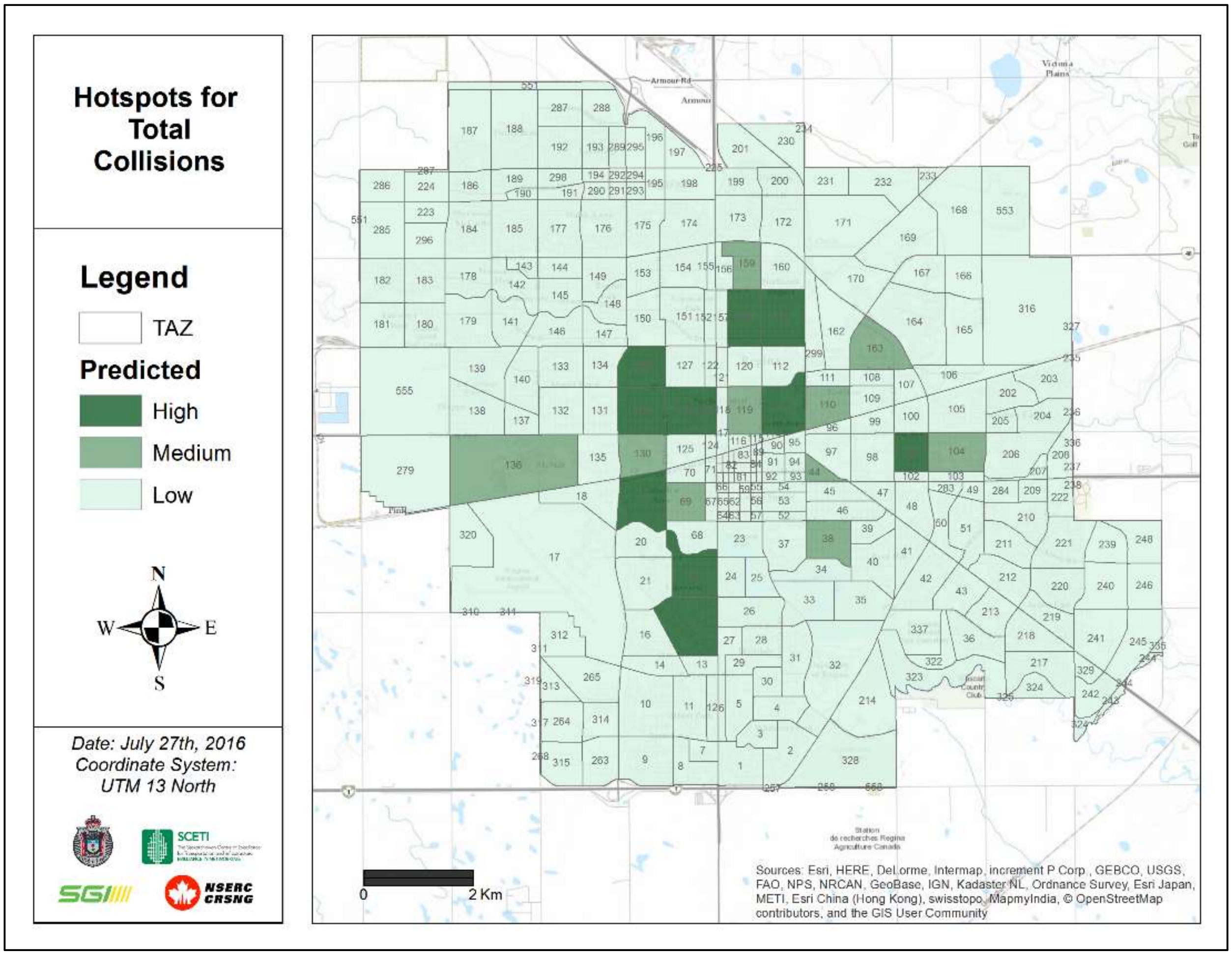
D11: Theft of Auto Crimes

Cumulative Residual Plots for Theft of Auto Crimes Top 6 Models

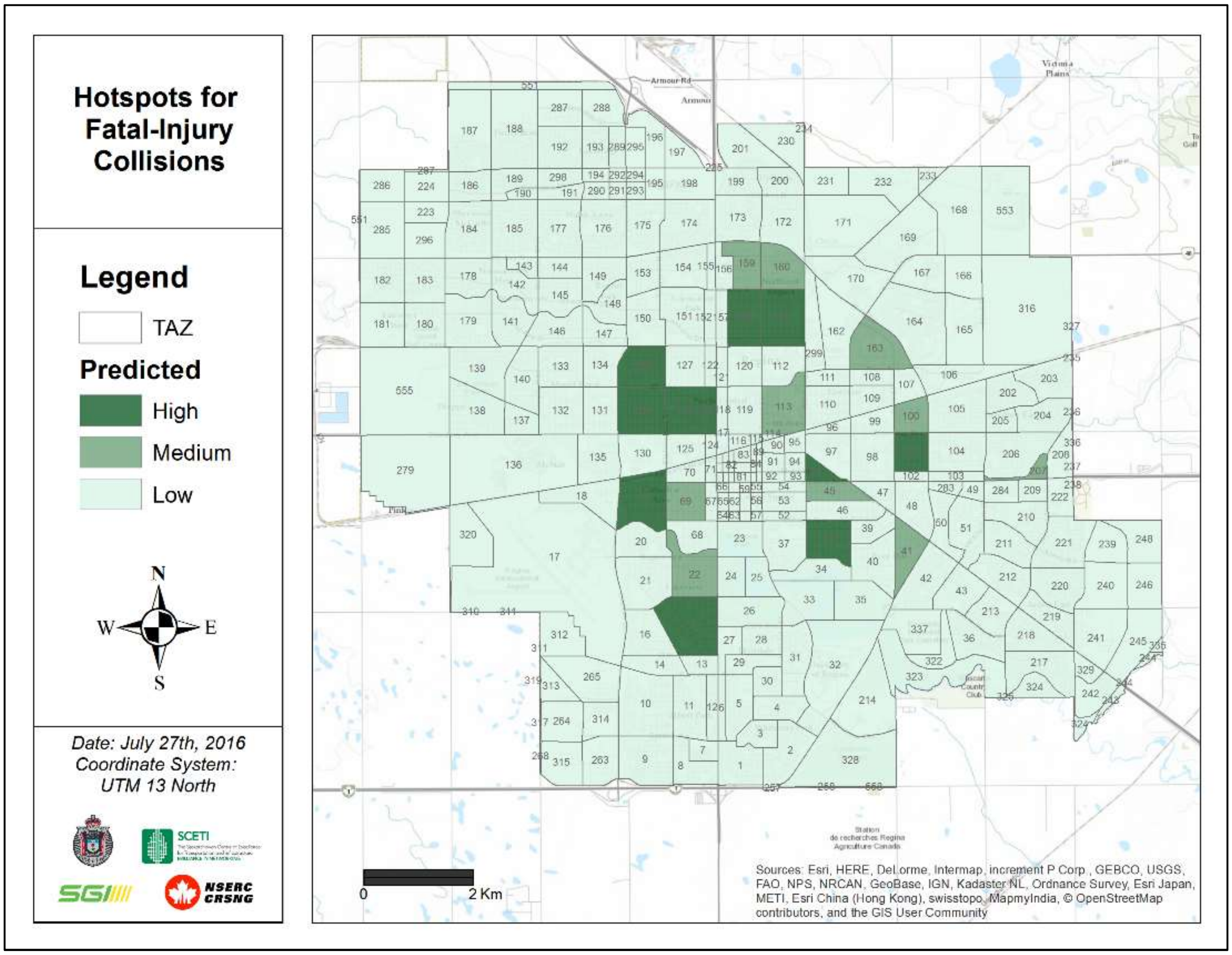


APPENDIX E: Prediction Hotspot Maps

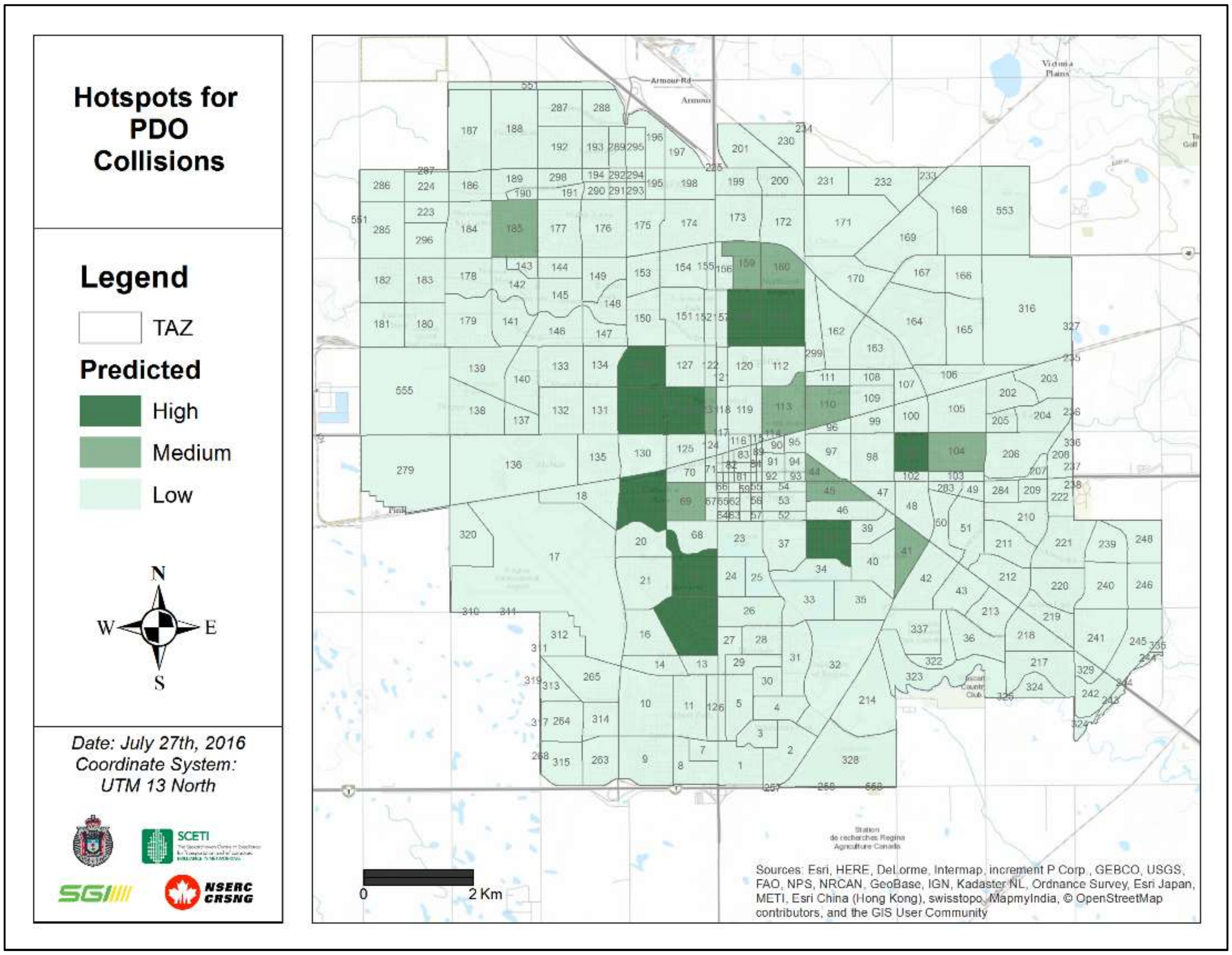
E1: Total Collisions



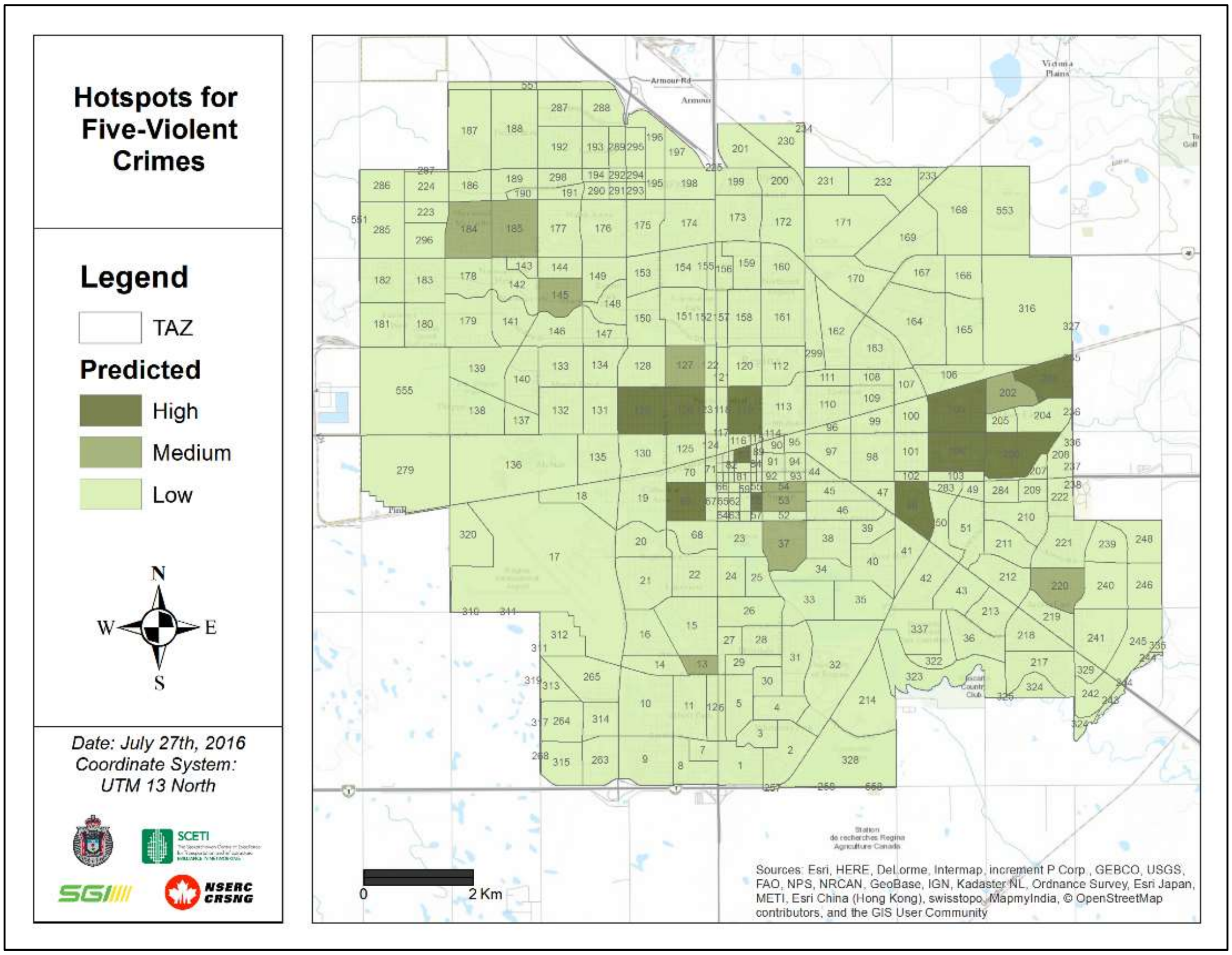
E2: Fatal-Injury Collisions



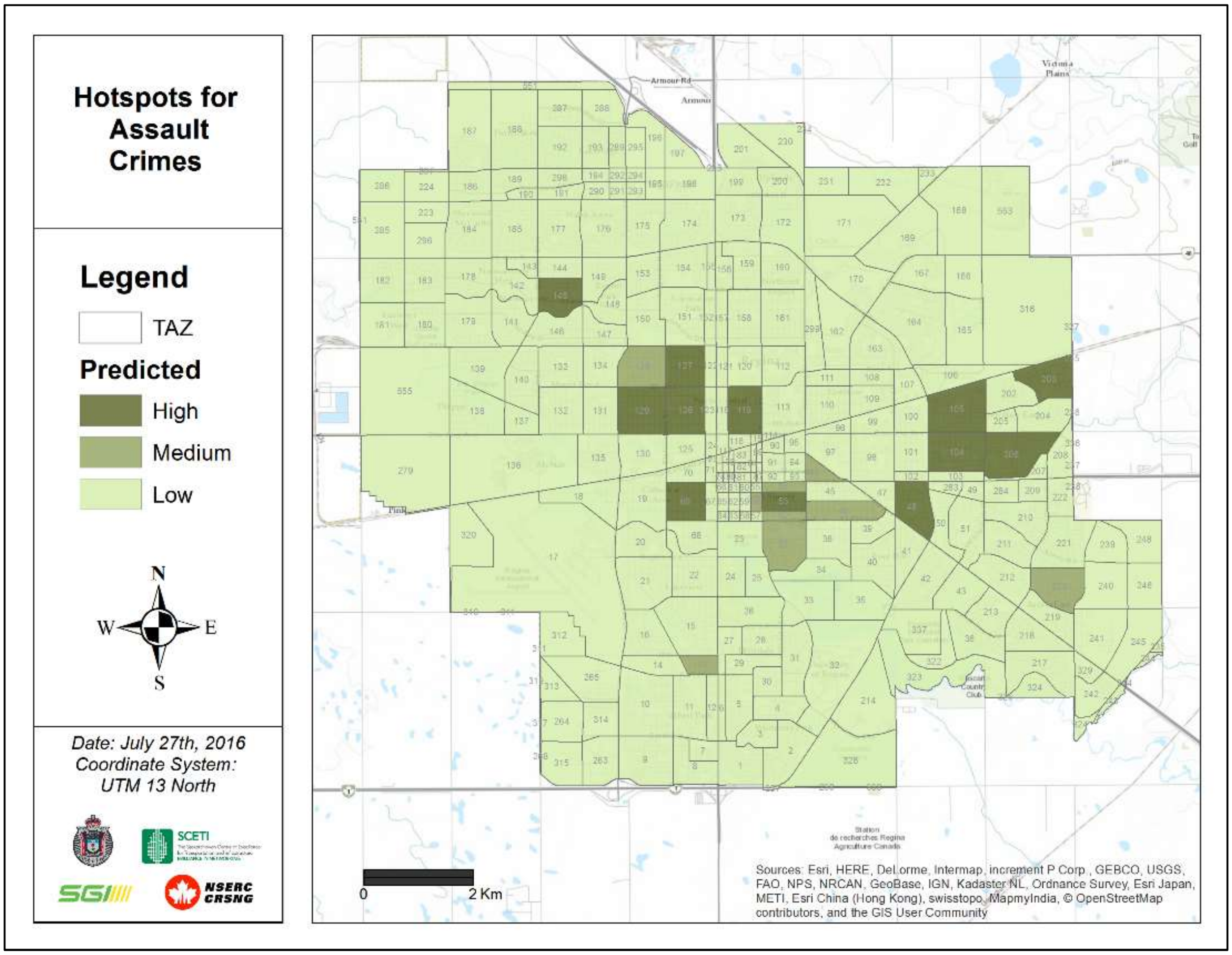
E3: Property Damage Only Collisions



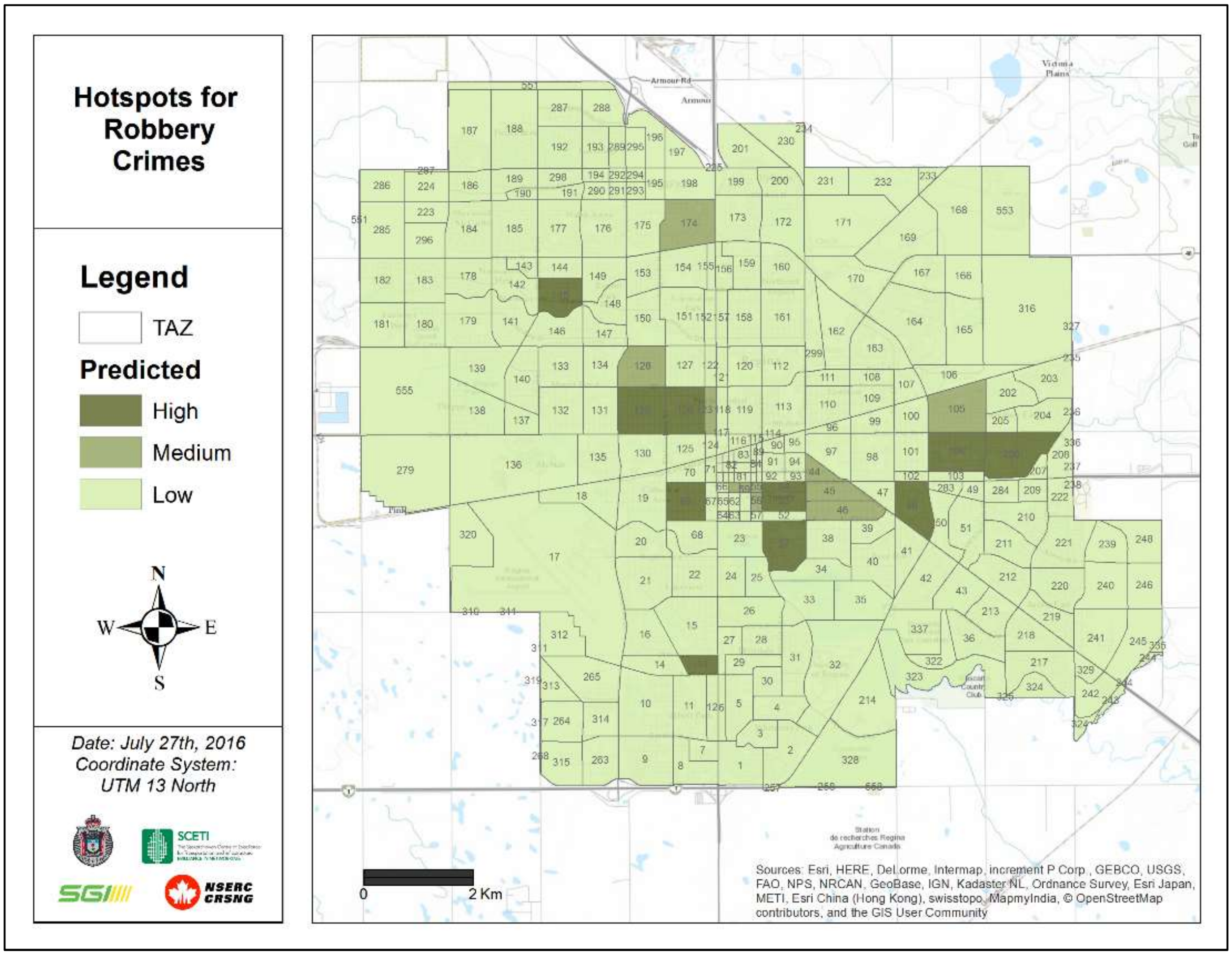
E4: Violent Crimes



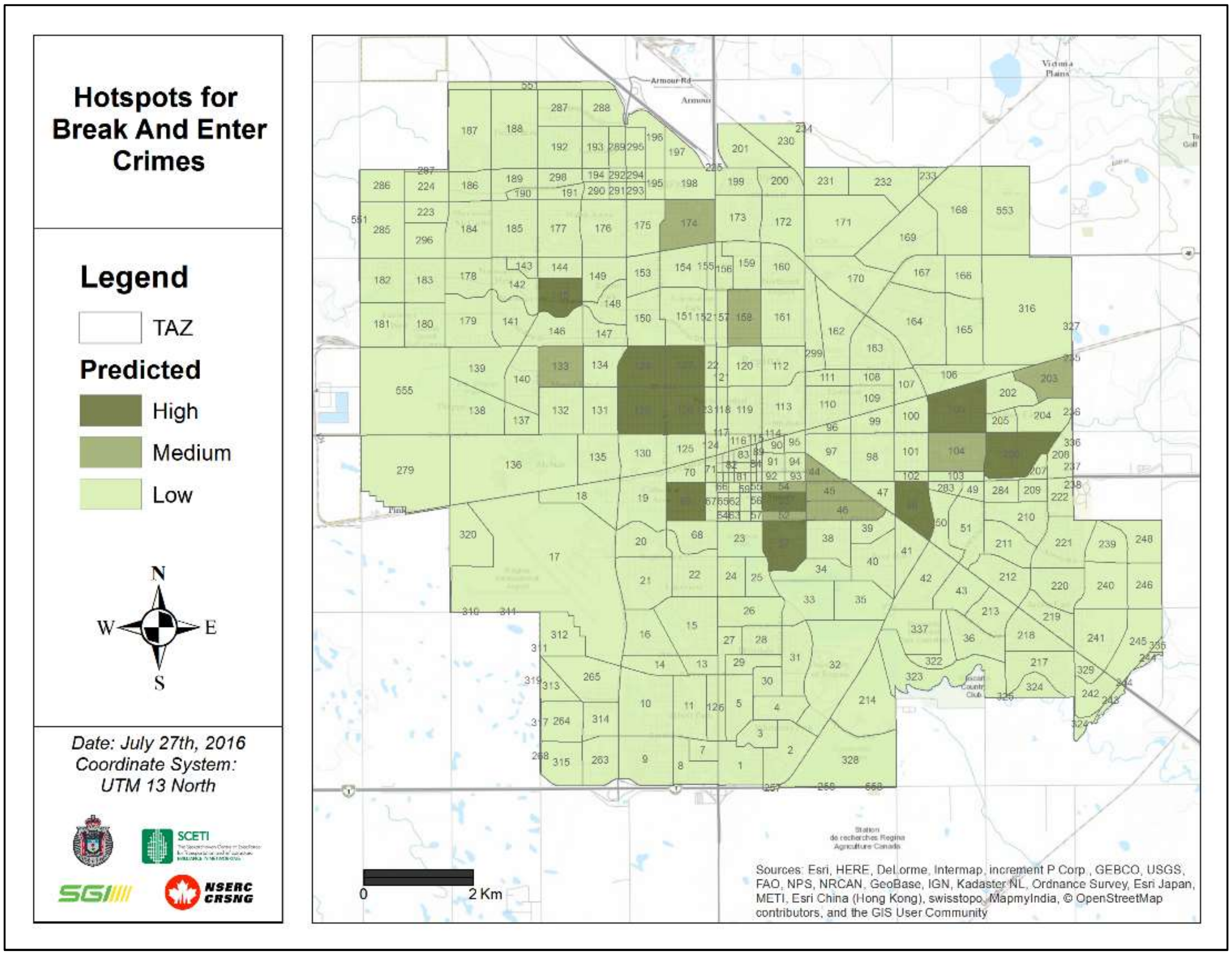
E5: Assault Crimes



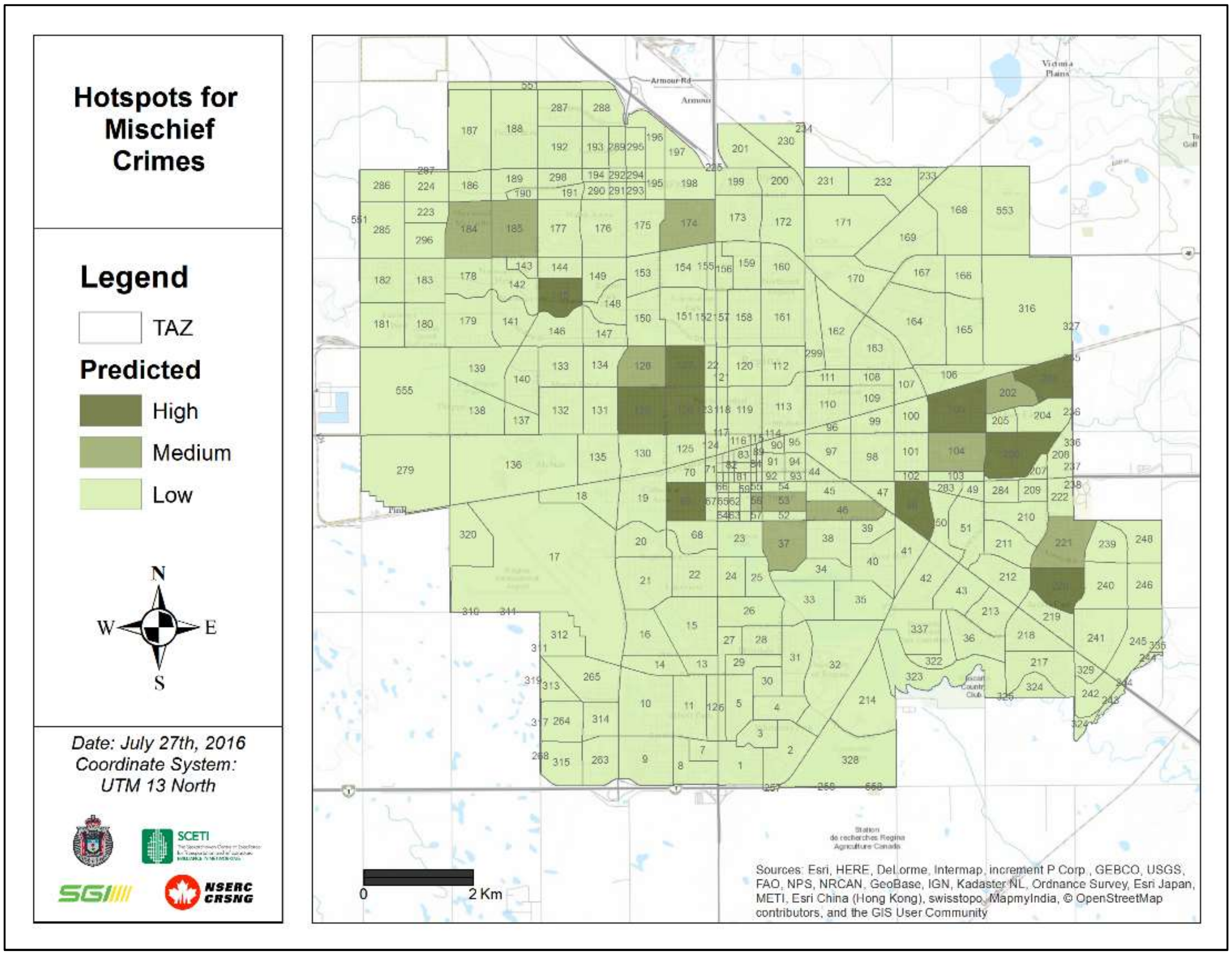
E6: Robbery Crimes



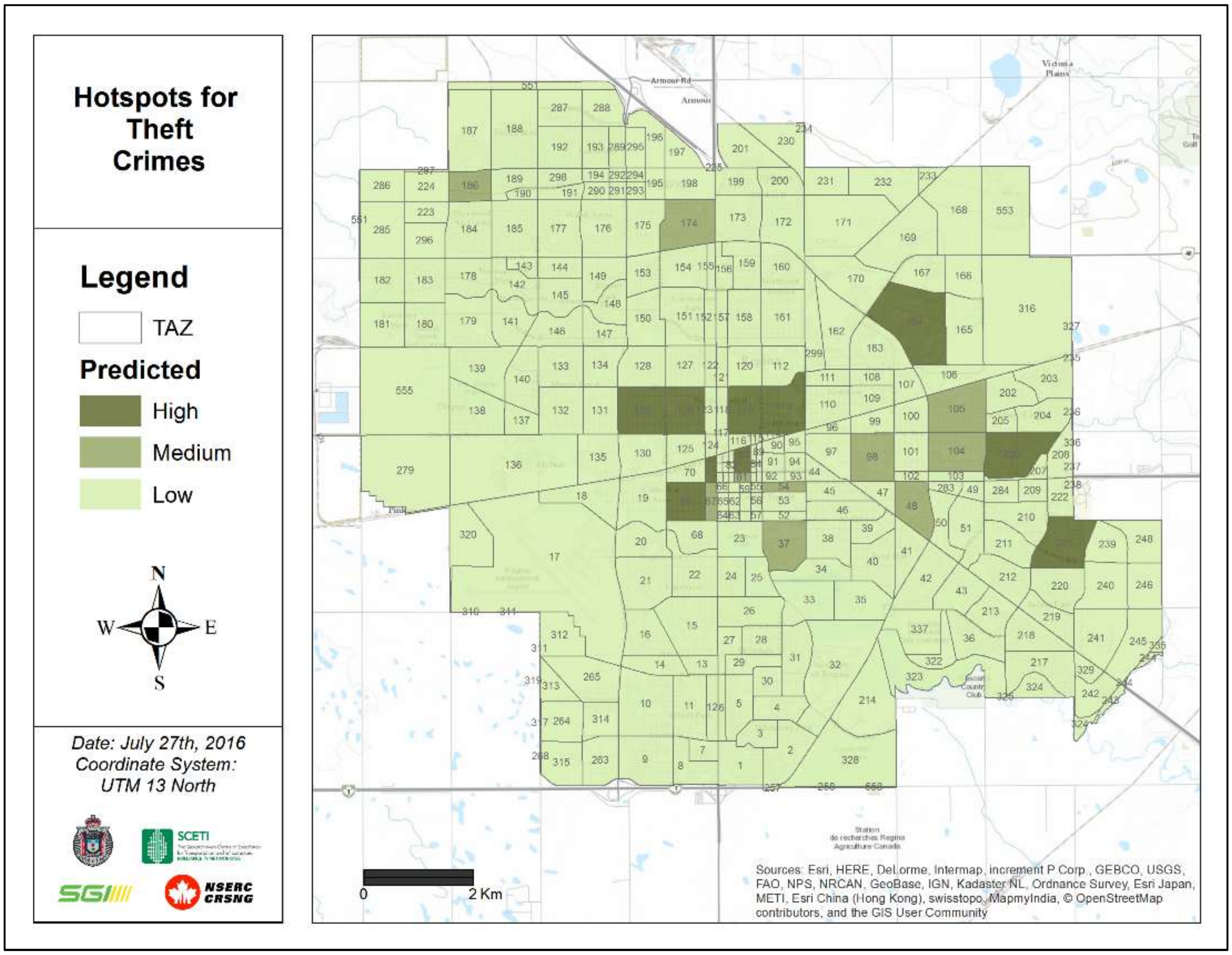
E7: Break and Enter Crimes



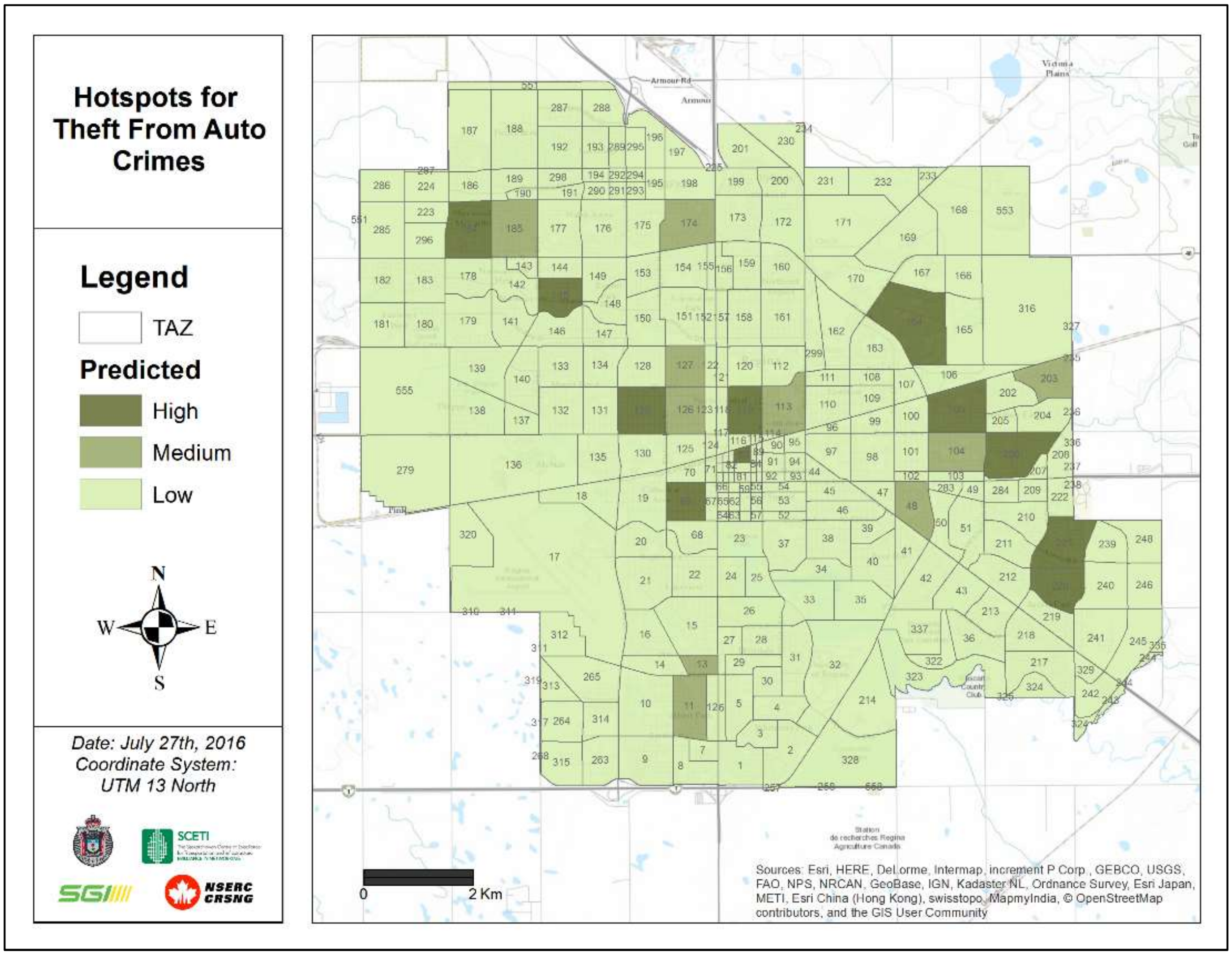
E8: Mischief Crimes



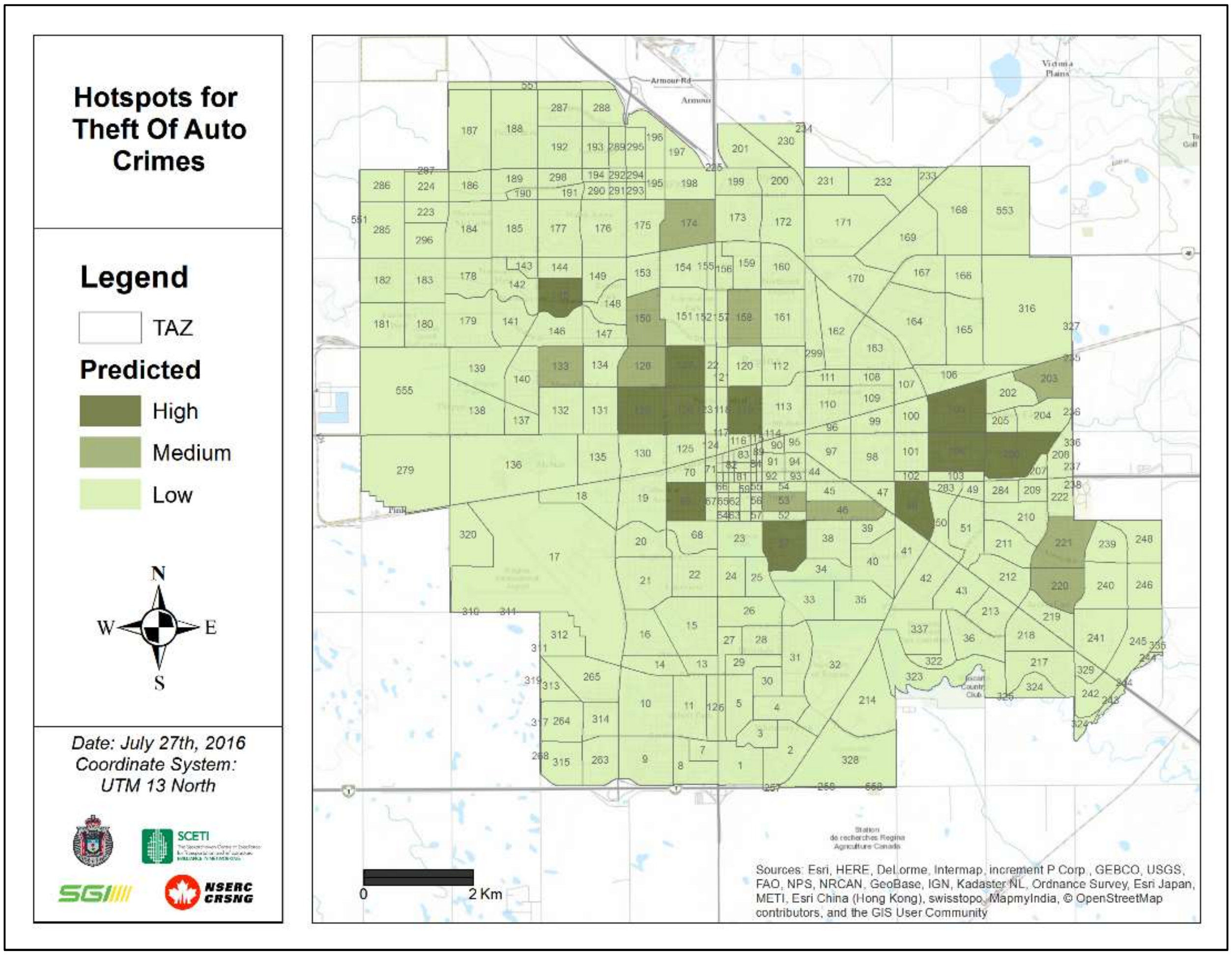
E9: Theft Crimes



E10: Theft from Auto Crimes

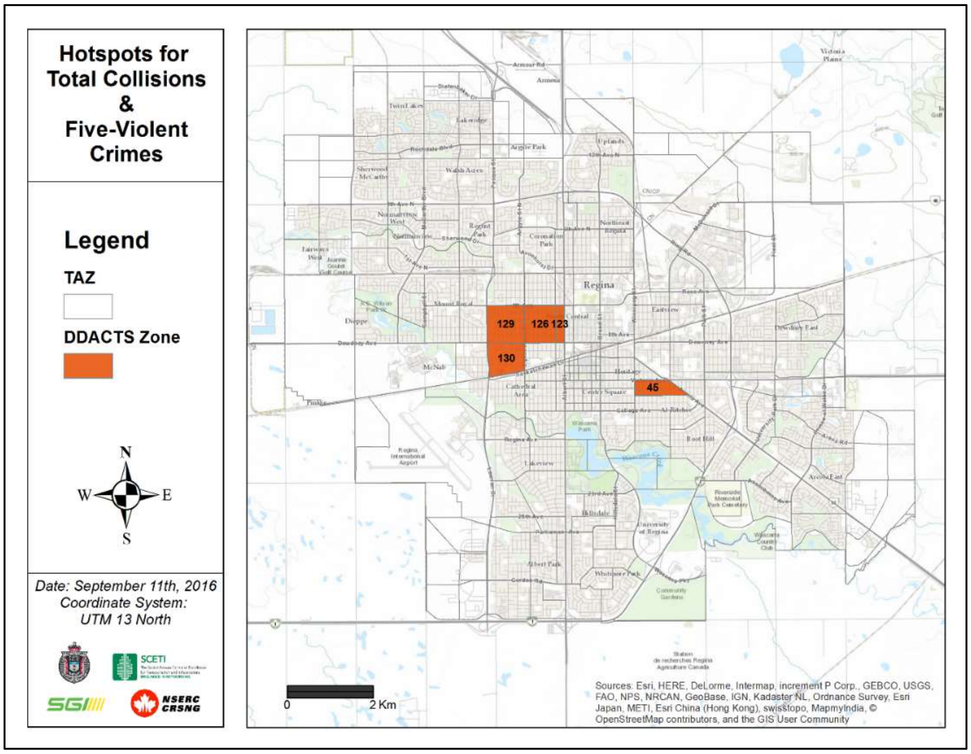


E11: Theft of Auto Crimes

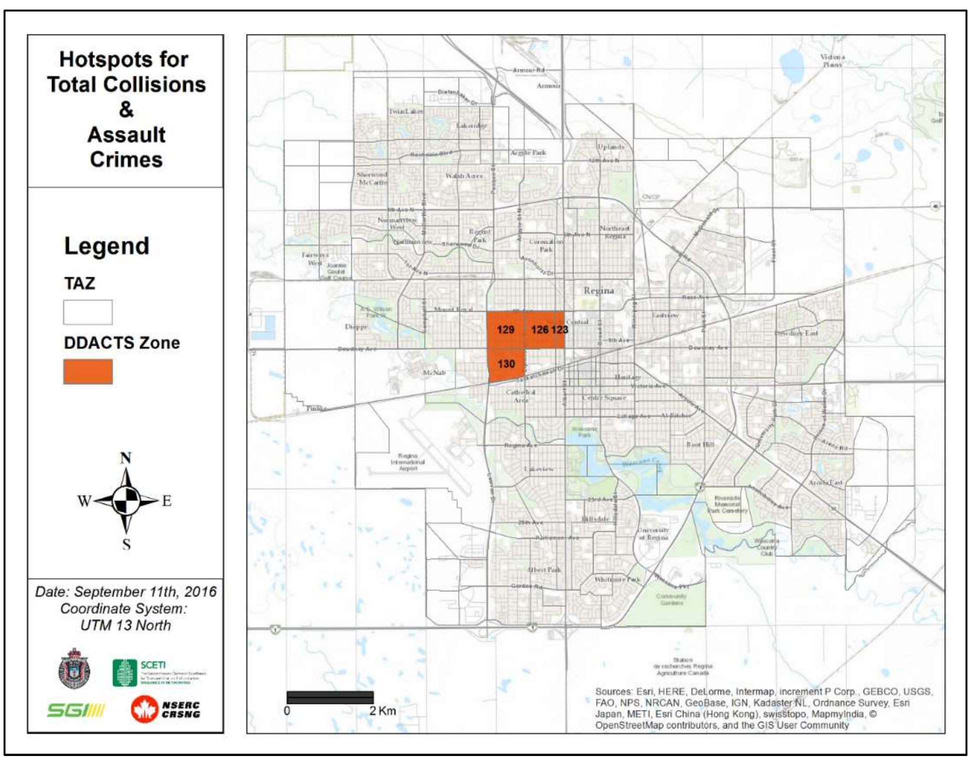


APPENDIX F: DDACTS Zone Maps

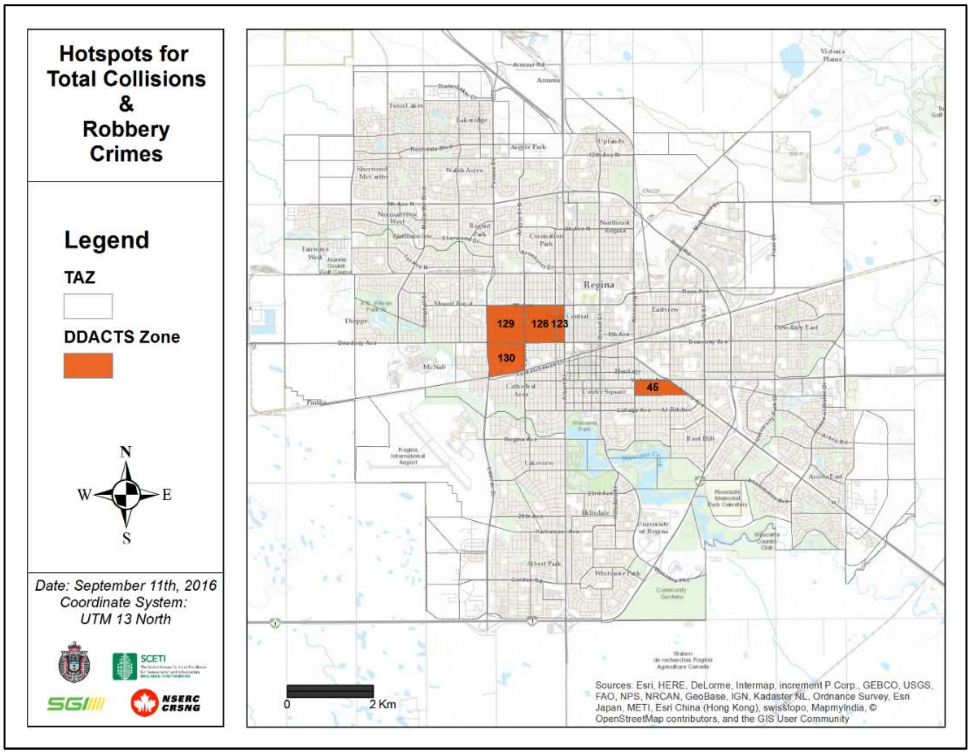
Total Collisions and Crimes: *DDACTS zone for total collisions and Violent crimes*



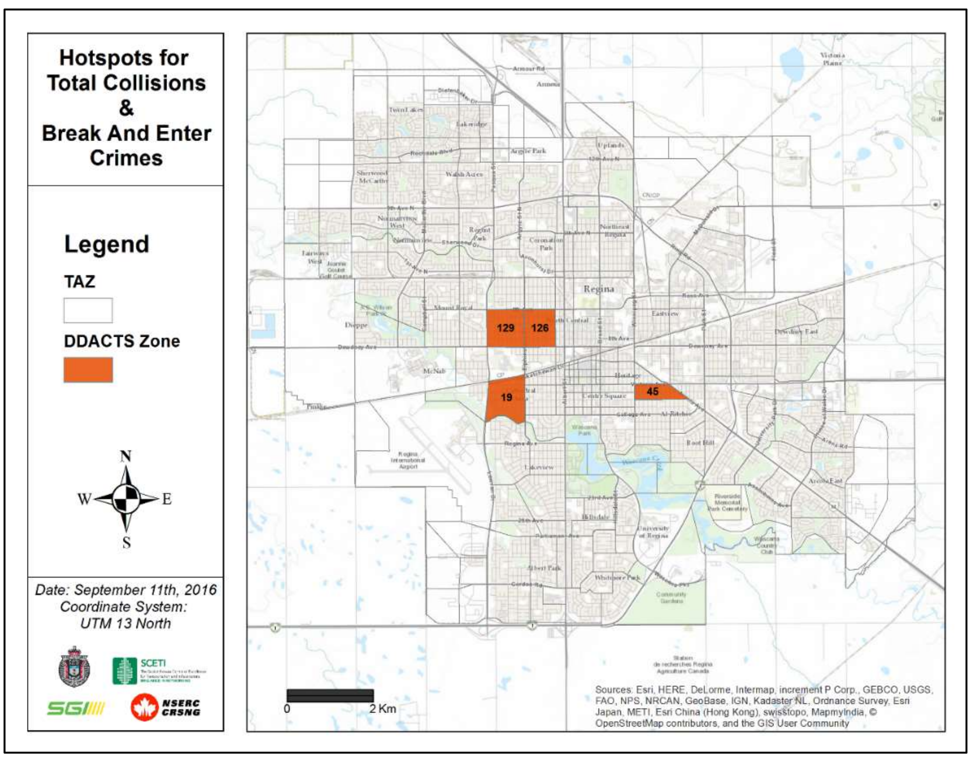
DDACTS zone for Total collisions and Assault crimes



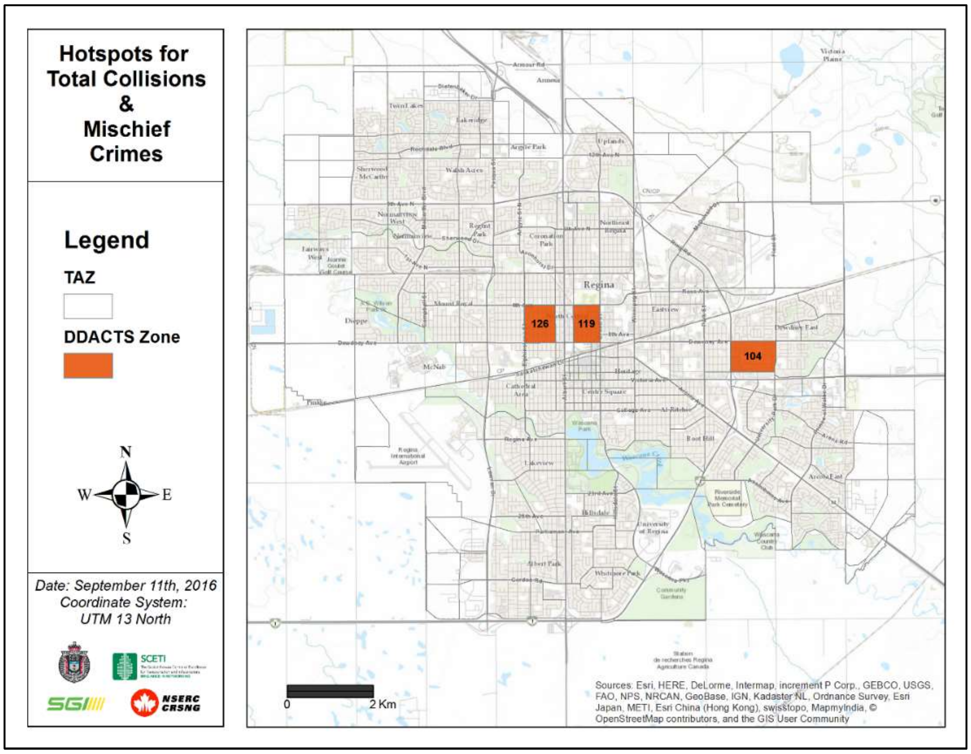
DDACTS zone for Total collisions and Robbery crimes



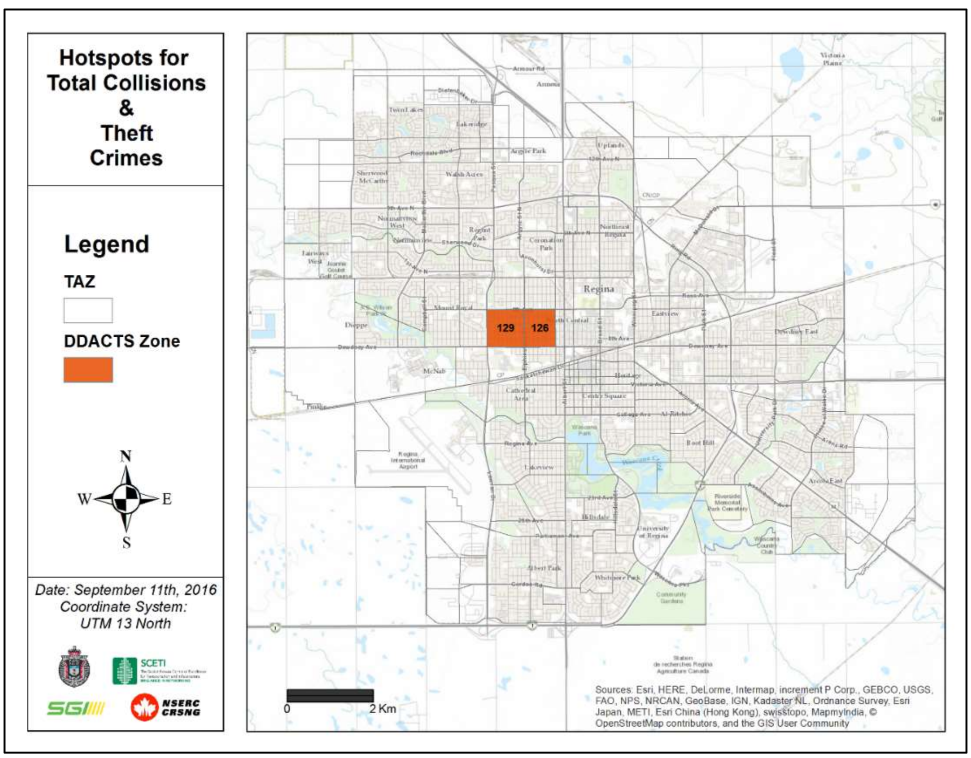
DDACTS zone for Total collisions and Break and Enter crimes



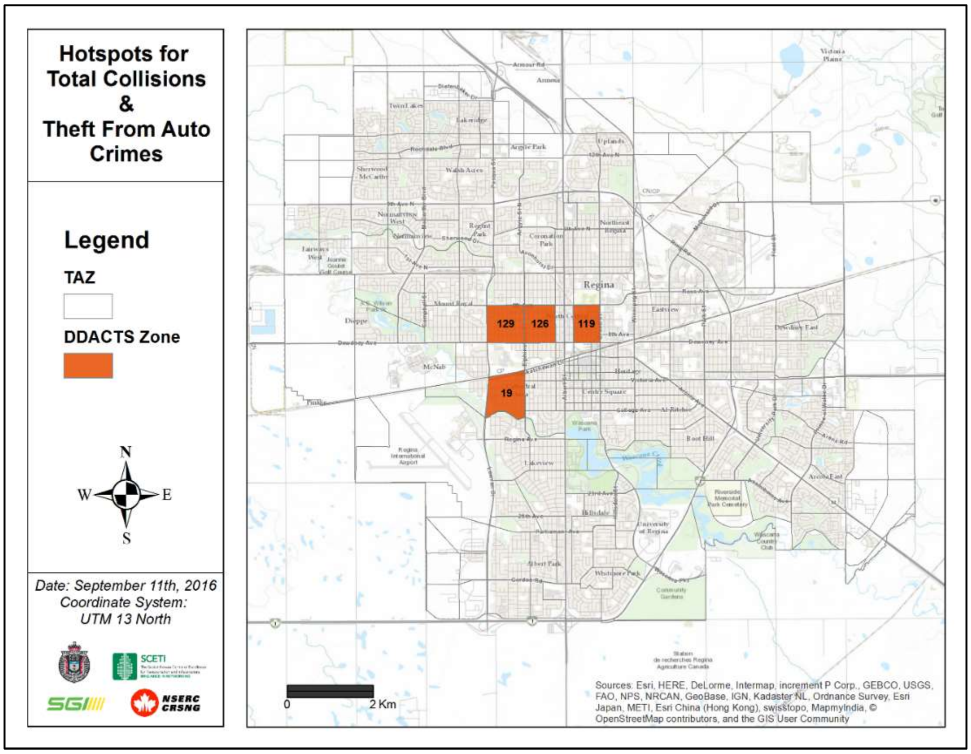
DDACTS zone for Total collisions and Mischief crimes



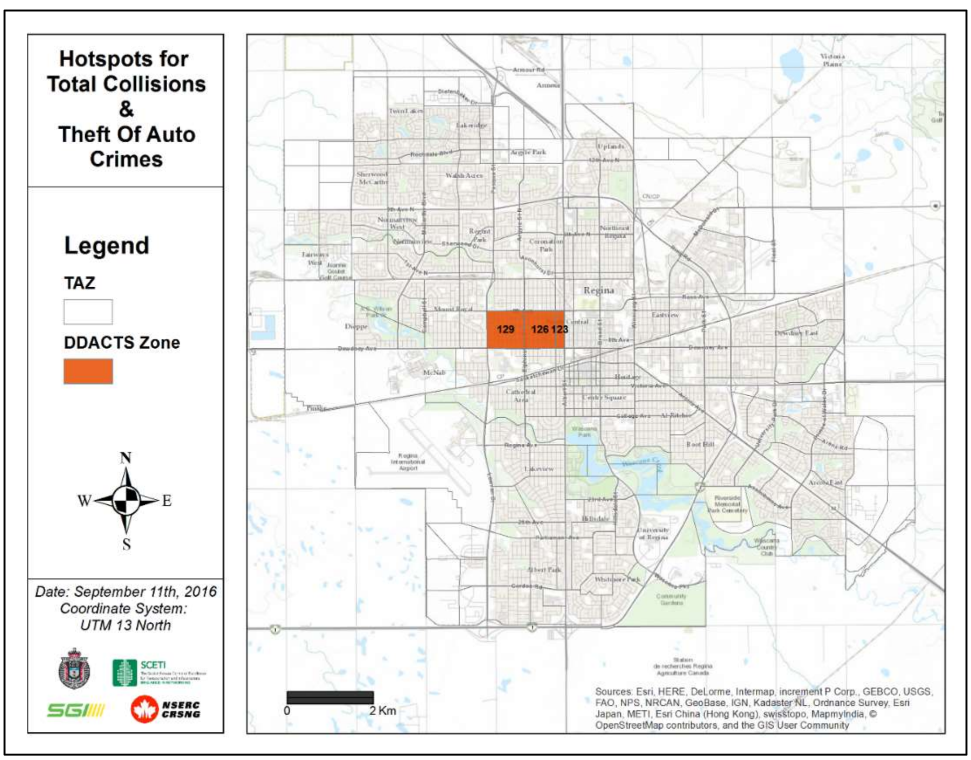
DDACTS zone for Total collisions and Theft crimes



DDACTS zone for Total collisions and Theft from Auto crimes

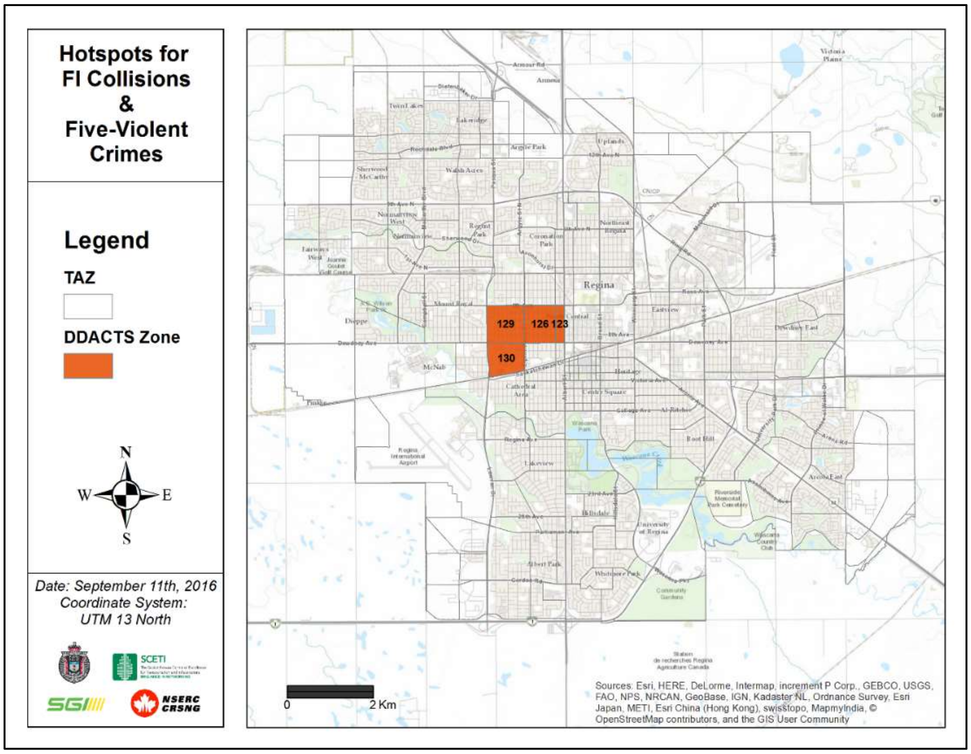


DDACTS zone for Total collisions and Theft of Auto crimes

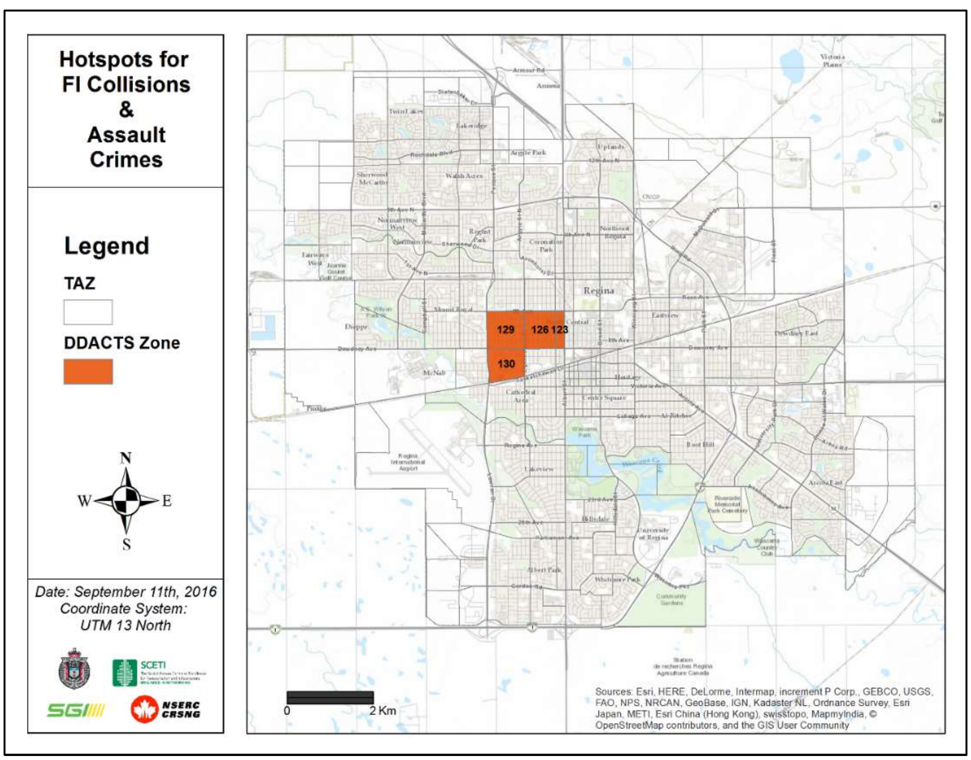


Fatal-Injury Collisions and Crimes

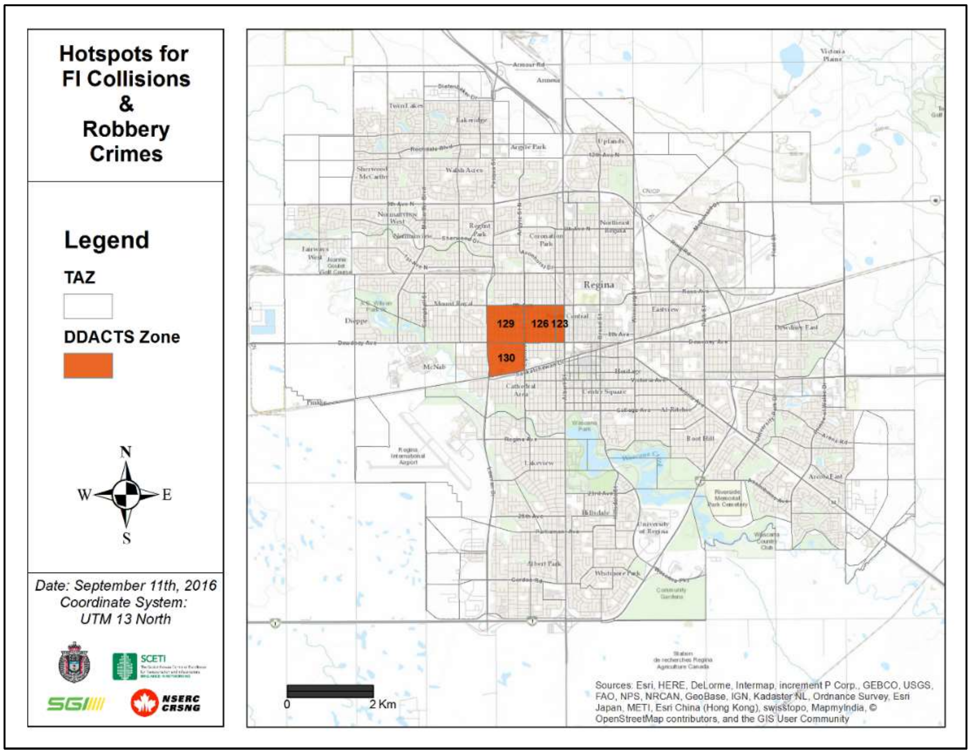
DDACTS zone for Fatal-Injury collisions and Violent crimes



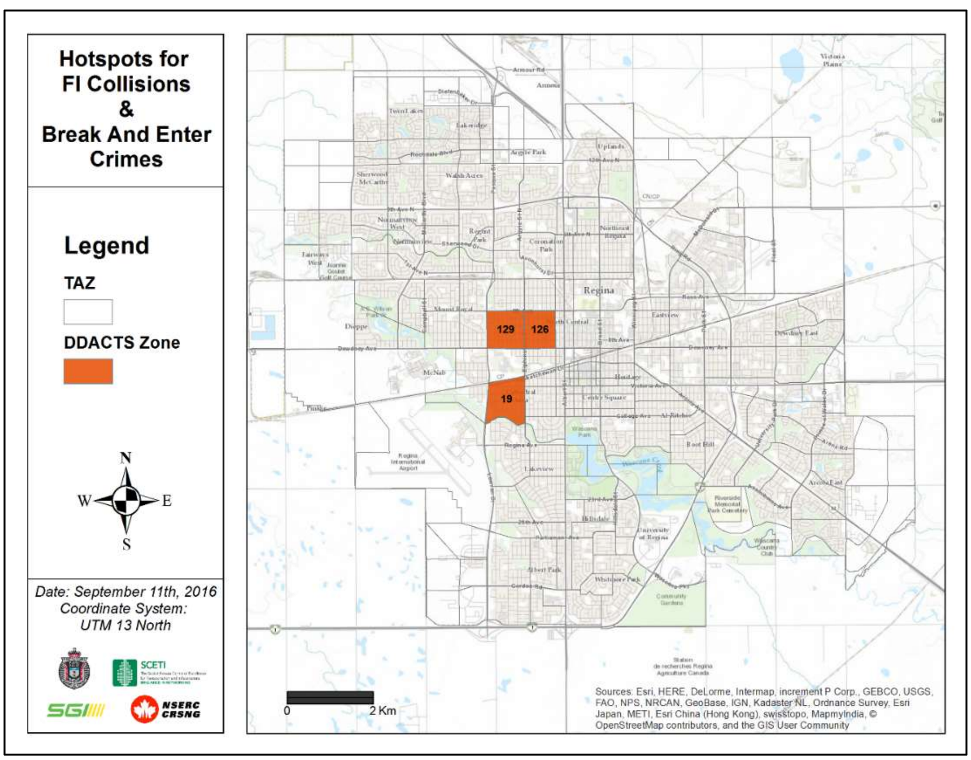
DDACTS zone for Fatal-Injury collisions and Assault crimes



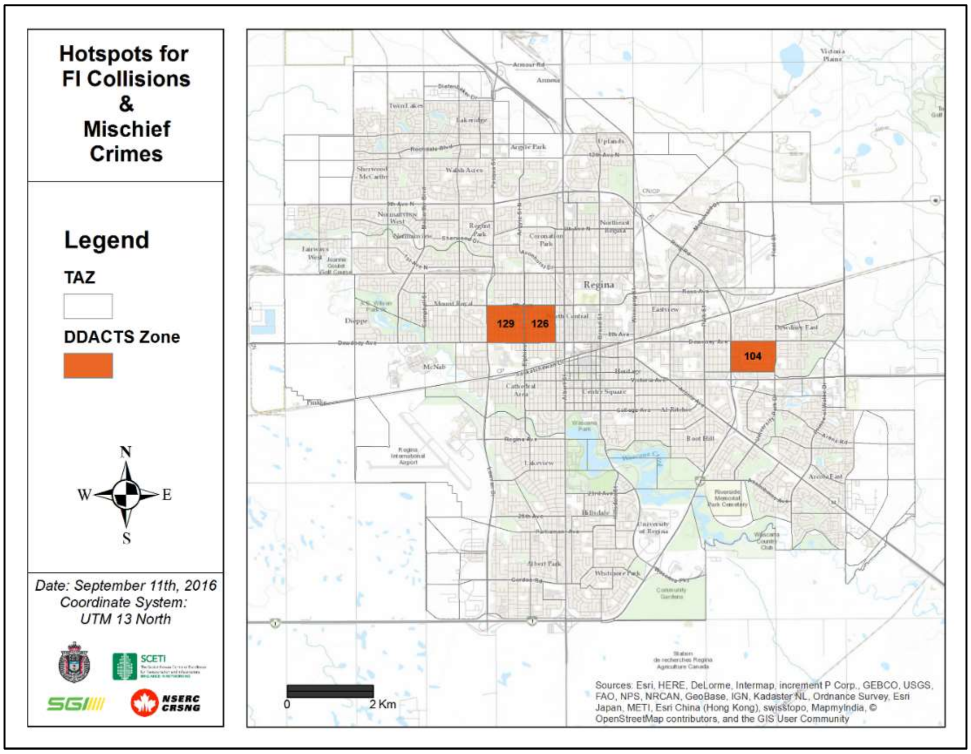
DDACTS zone for Fatal-Injury collisions and Robbery crimes



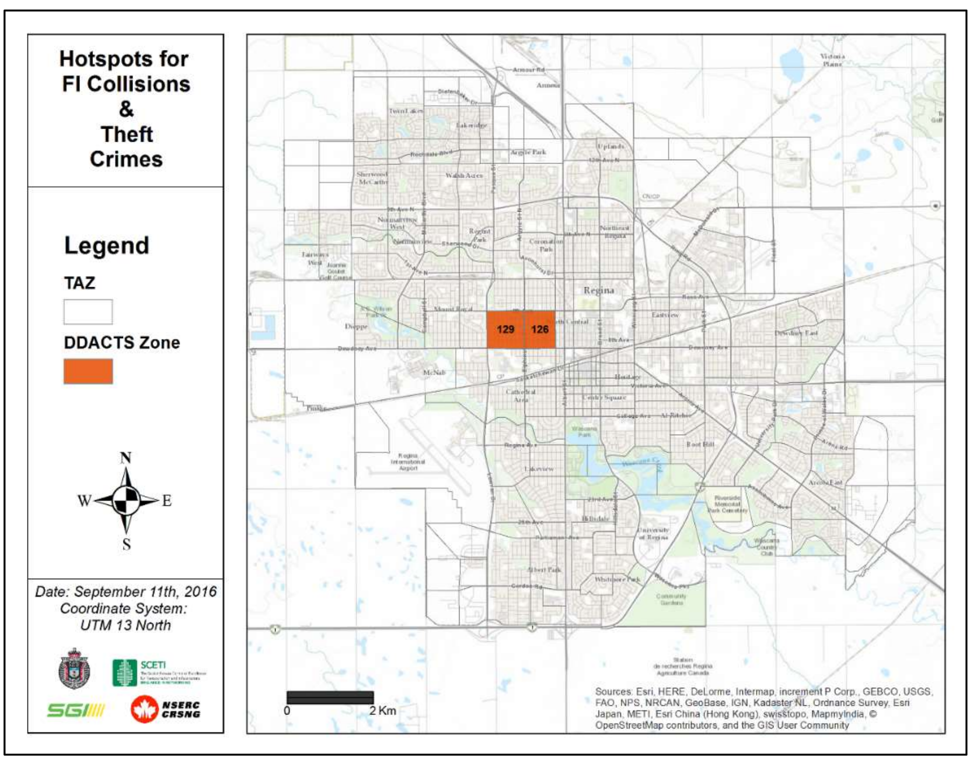
DDACTS zone for Fatal-Injury collisions and Break and Enter crimes



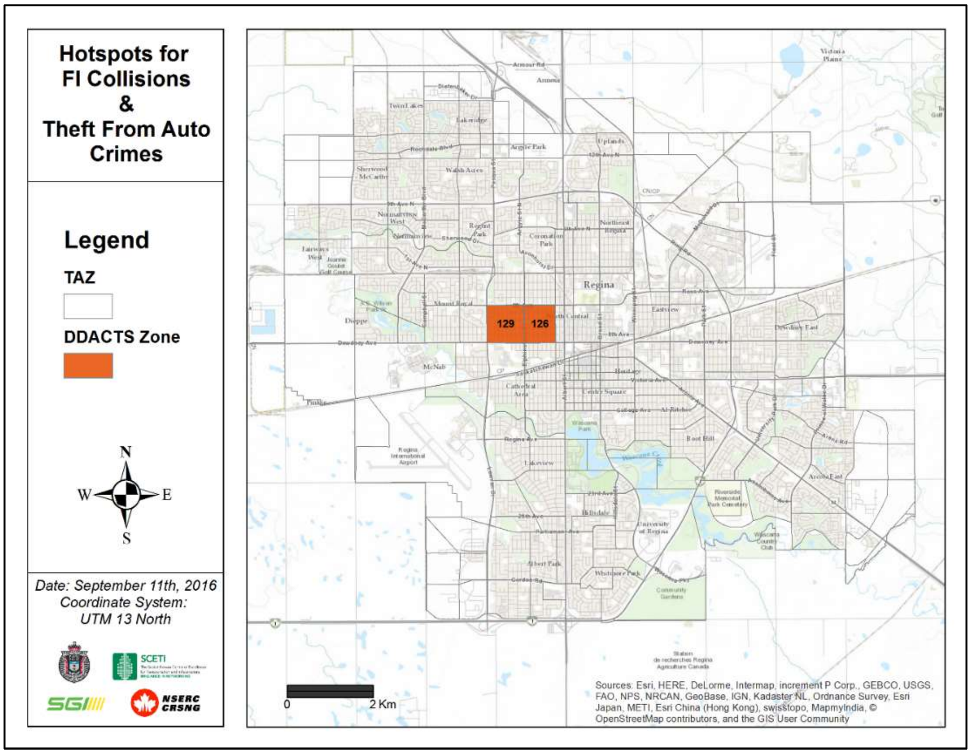
DDACTS zone for Fatal-Injury collisions and Mischief crimes



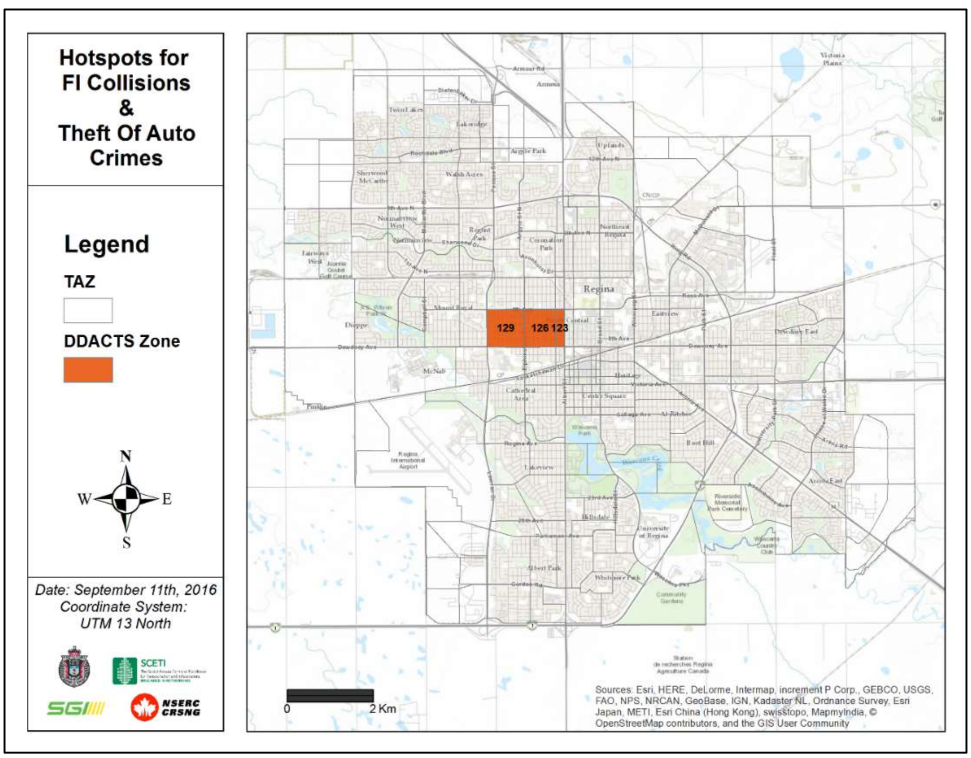
DDACTS zone for Fatal-Injury collisions and Theft crimes



DDACTS zone for Fatal-Injury collisions and Theft from Auto crimes

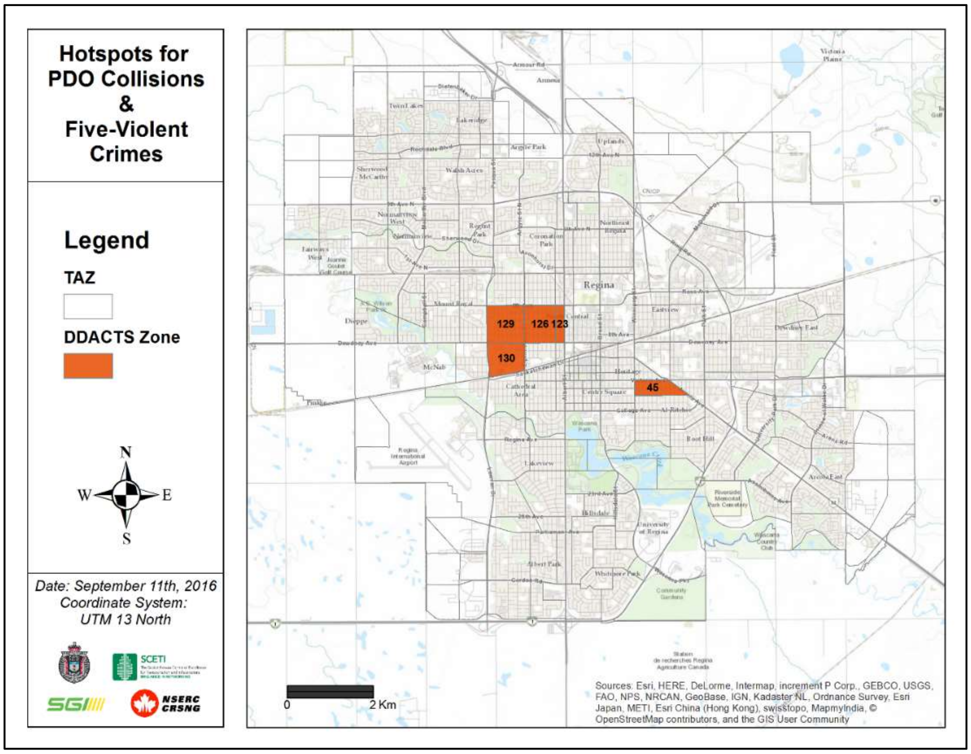


DDACTS zone for Fatal-Injury collisions and Theft from Auto crimes

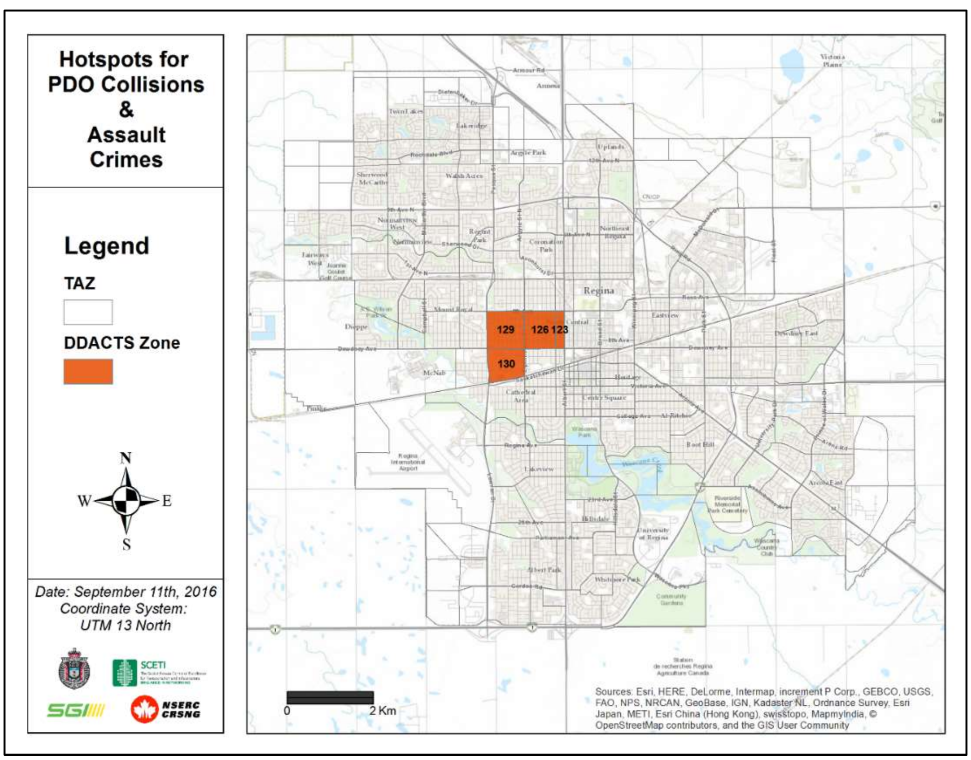


Property Damage Only Collisions and Crimes

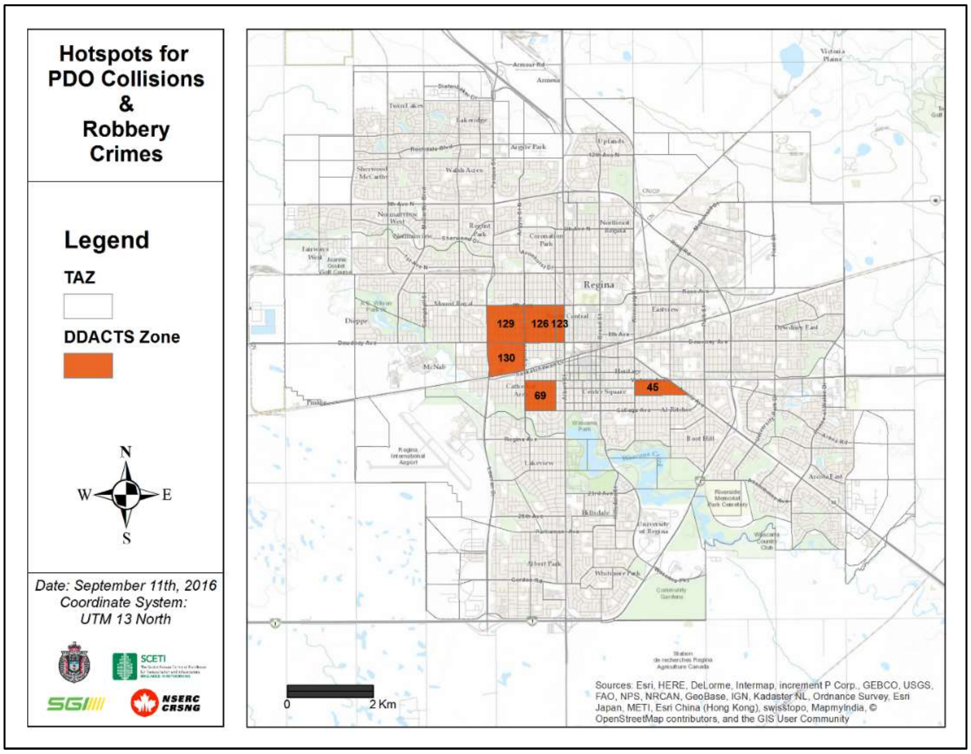
DDACTS zone for Property Damage Only collisions and Violent crimes



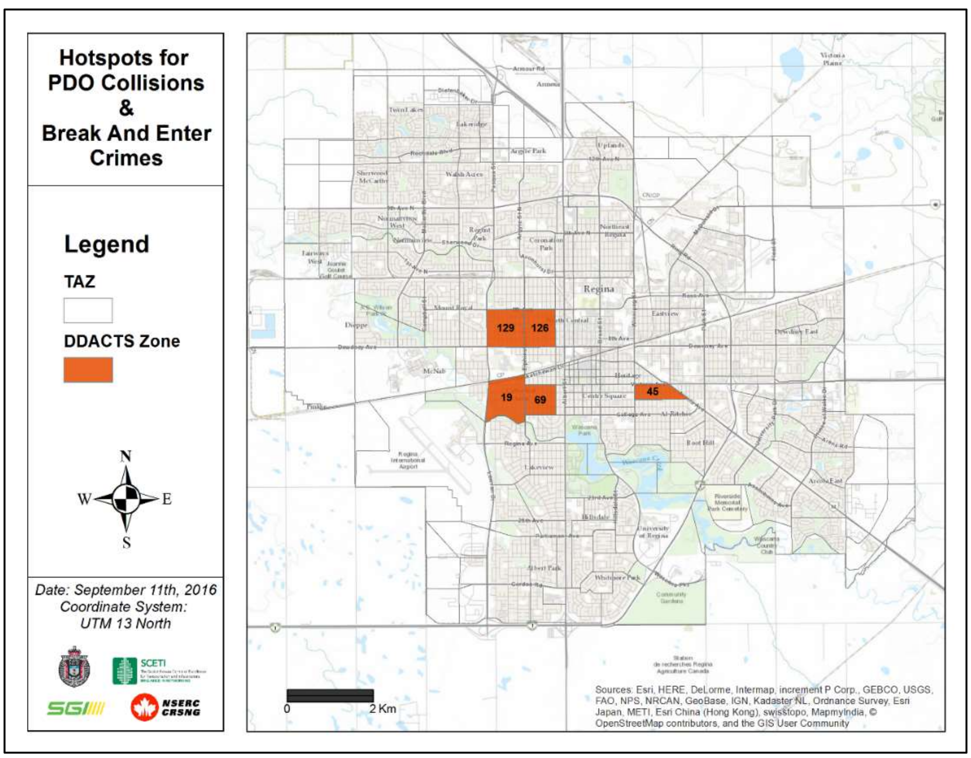
DDACTS zone for Property Damage Only collisions and Assault crimes



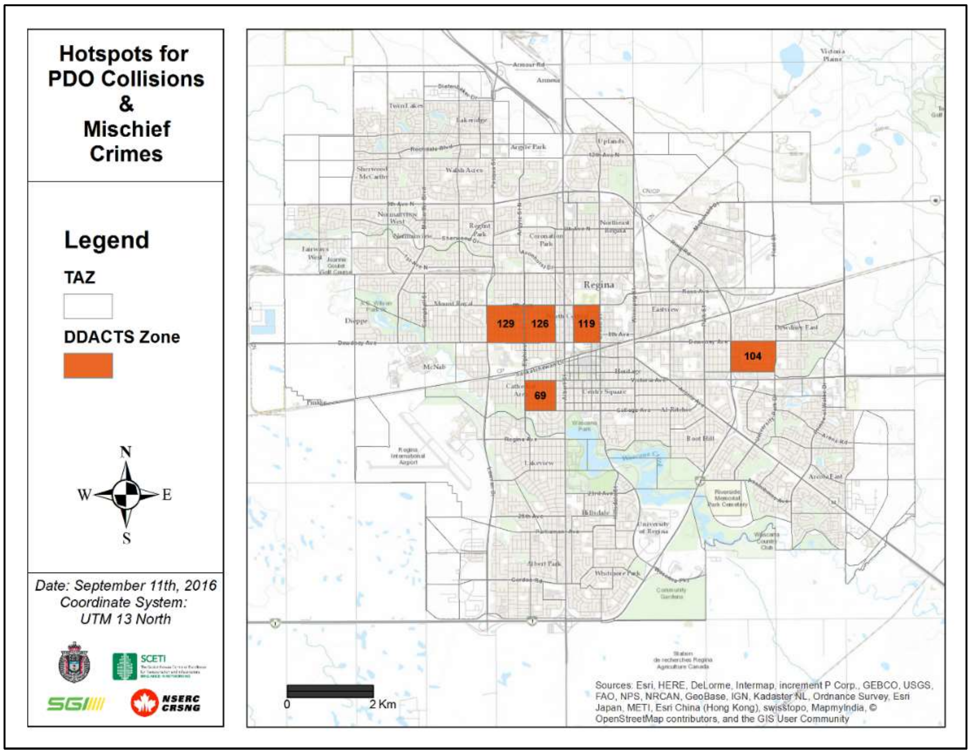
DDACTS zone for Property Damage Only collisions and Robbery crimes



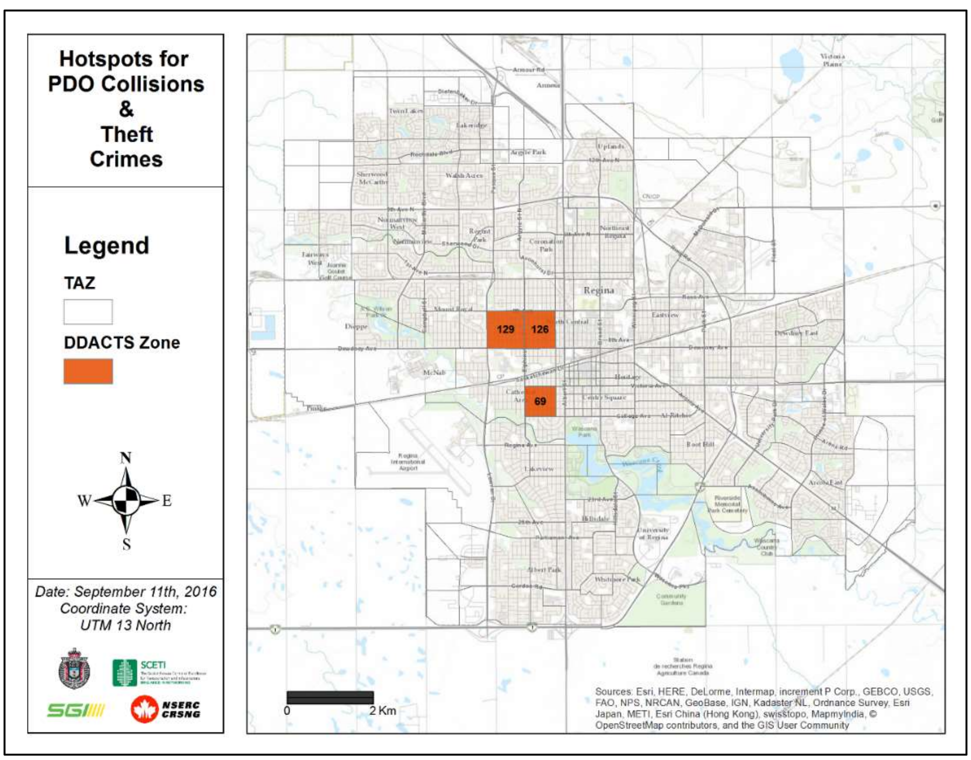
DDACTS zone for Property Damage Only collisions and Break and Enter crimes



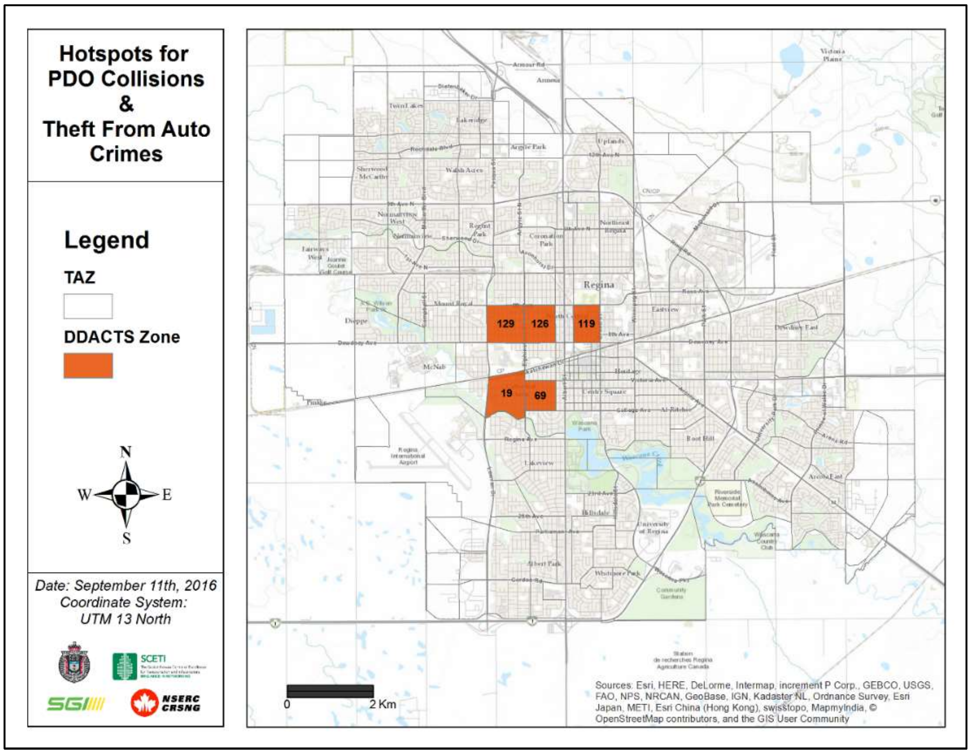
DDACTS zone for Property Damage Only collisions and Mischief crimes



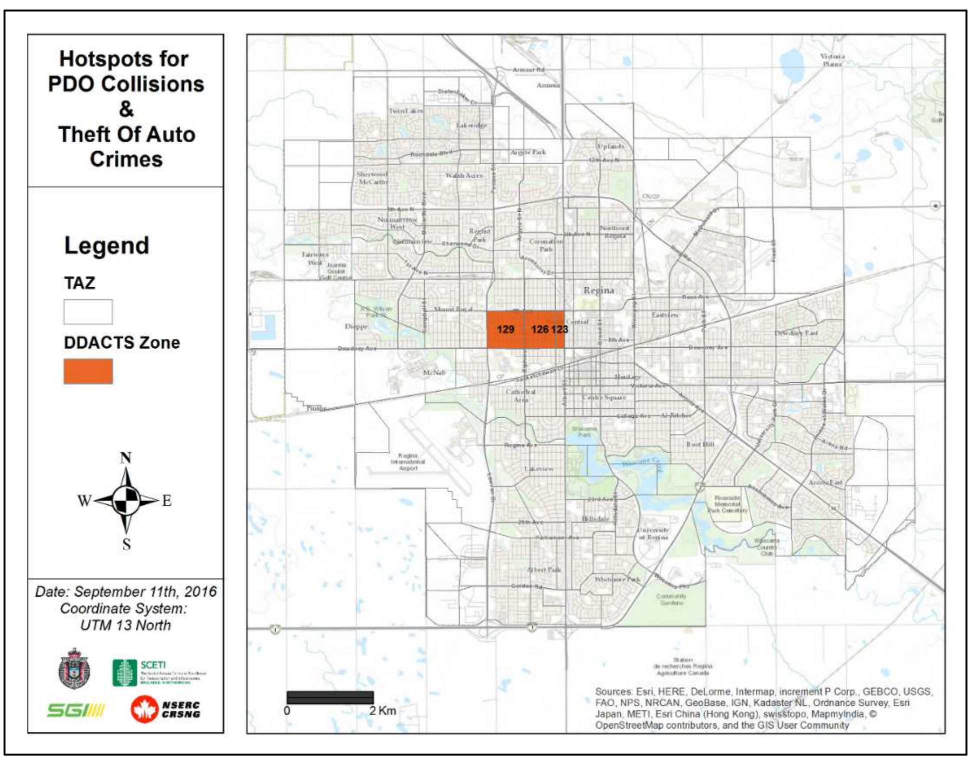
DDACTS zone for Property Damage Only collisions and Theft crimes



DDACTS zone for Property Damage Only collisions and Theft from Auto crimes



DDACTS zone for Property Damage Only collisions and Theft of Auto crimes



APPENDIX G: Summary of TAZ Data Excluded from Analysis.

TAZ Number	Total Population	TAZ Area (m^2)	VKMT	Total Collisions	FI Collisions	PDO Collisions
195	250	201795.0118	0	0	0	0
196	481	415624.1481	0	0	0	0
197	490	334079.2136	0	0	0	0
114	9	37233.51343	0	0	0	0
171	0	1561297.764	0	0	0	0
224	215	453622.4194	0	0	0	0
311	1	6568.892387	0	0	0	0
319	1	12760.353	0	0	0	0
553	47	1428333.771	0	0	0	0
233	0	137003.4318	0	0	0	0
234	0	910.3518375	0	0	0	0
235	0	1073.991747	0	0	0	0
238	2	3531.764955	0	0	0	0
225	49	1.647867426	0	0	0	0
231	5	416957.9111	0	0	0	0
232	0	743722.6466	0	0	0	0
258	2	14990.76723	0	0	0	0
268	1	13029.55645	0	0	0	0
239	0	471701.6205	0	0	0	0
240	583	698774.4534	0	0	0	0
243	50	88281.29868	0	0	0	0
244	40	58816.73057	0	0	0	0
246	0	520543.7142	0	0	0	0
248	0	556501.8523	0	0	0	0
257	4	21684.55214	0	0	0	0
295	335	272421.2166	0	0	0	0
310	1	5457.427518	0	0	0	0
322	28	335815.8777	0	0	0	0
325	1	348.8932586	0	0	0	0
335	17	34234.12923	0	0	0	0
276	1	23707.07672	0	0	0	0
554	0	17449.97052	0	0	0	0
558	108	9057.954794	0	0	0	0
285	385	825114.9947	0	0	0	0
297	23	31961.91766	0	0	0	0
286	232	479777.6905	0	0	0	0
327	0	6995.158488	0	0	0	0

APPENDIX H: Regression Model Results for Top Candidate Models

Results from regression analysis for the top 10 Collision Prediction Models for each severity type and the top 6 crime prediction models for each crime occurrence type are summarized in Tables. Mathematical equations representing each model are also presented.

H.1 Total Collisions Models

Model 1

$$TOT = N \times \exp(-1.98) \times$$

$$\exp\left(\begin{array}{l} (\log VKMT \times 0.475) + (URBAN\_HOLDING\_RESIDENTIAL\_AREA\_PROP \times -2.59) + (POPULATION\_DENSITY \times 1.42 \times 10^{-4}) + \\ (FOUR\_LEG\_INTERSECTIONS \times 0.0435) + (THREE\_LEG\_INTERSECTIONS \times -5.97 \times 10^{-3}) + (POPULATION\_18TO24 \times 3.41 \times 10^{-3}) + \\ (POPULATION\_45TO64 \times -2.15 \times 10^{-3}) \end{array} \right)$$

Model 1					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-1.98E+00	5.51E-01	-3.59069	< 0.001	Dispersion Parameter = 1.8227 Standard Error = 0.185 Log-likelihood = -2449.502
logVKMT	4.75E-01	5.81E-02	8.18662	< 0.001	
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	-2.59E+00	2.68E-01	-9.67222	< 0.001	
POPULATION_DENSITY	1.42E-04	3.71E-05	3.83632	< 0.001	
FOUR_LEG_INTERSECTIONS	4.35E-02	1.02E-02	4.28133	< 0.001	
THREE_LEG_INTERSECTIONS	-5.97E-03	7.75E-03	-0.76952	0.4416	
POPULATION_18TO24	3.41E-03	2.24E-03	1.52011	0.1285	
POPULATION_45TO64	-2.15E-03	8.55E-04	-2.51312	0.0120	

Model 2

$$TOT = N \times \exp(-2.01) \times$$

$$\exp\begin{pmatrix} (\log VKMT \times 0.474) + (FOUR\_LEG\_INTERSECTIONS \times 0.0396) + (INTERSECTION\_DENSITY \times 7.55 \times 10^{3}) + \\ (SEGMENT\_80KMHR \times -8.47 \times 10^{-4}) + (THREE\_LEG\_INTERSECTIONS \times -0.0122) + (LOW\_DENSITY\_RESIDENTIAL\_ \\ AREA\_PROP \times -0.417) + (POPULATION\_18TO24 \times 5.07 \times 10^{-3}) + (POPULATION\_45TO64 \times -1.74 \times 10^{-3}) \end{pmatrix}$$

Model 2					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-2.01E+00	5.60E-01	-3.59858	< 0.001	Dispersion Parameter = 1.4282 Standard Error = 0.142 Log-likelihood = -2508.847
logVKMT	4.74E-01	6.06E-02	7.82689	< 0.001	
FOUR_LEG_INTERSECTIONS	3.96E-02	1.19E-02	3.31359	< 0.001	
INTERSECTION_DENSITY	7.55E+03	2.12E+03	3.56456	< 0.001	
SEGMENT_80KMHR	-8.47E-04	1.46E-04	-5.78274	< 0.001	
THREE_LEG_INTERSECTIONS	-1.22E-02	8.45E-03	-1.44637	0.1481	
LOW_DENSITY_RESIDENTIAL_AREA_PROP	-4.17E-01	2.63E-01	-1.58458	0.1131	
POPULATION_18TO24	5.07E-03	2.55E-03	1.98572	0.0471	
POPULATION_45TO64	-1.74E-03	9.56E-04	-1.81577	0.0694	

Model 3

$$TOT = N \times \exp(-1.77) \times$$

$$\exp\begin{pmatrix} (\log VKMT \times 0.444) + (FOUR\_LEG\_INTERSECTIONS \times 0.0451) + (INTERSECTION\_DENSITY \times 7.53 \times 10^{3}) + \\ (SEGMENT\_80KMHR \times -8.64 \times 10^{-4}) + (LOW\_DENSITY\_RESIDENTIAL\_AREA \times -0.506) \end{pmatrix}$$

Model 3					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-1.77E+00	5.20E-01	-3.41297	< 0.001	Dispersion Parameter = 1.3895 Standard Error = 0.137 Log-likelihood = -2514.520
logVKMT	4.44E-01	5.52E-02	8.05022	< 0.001	
FOUR_LEG_INTERSECTIONS	4.51E-02	1.10E-02	4.09201	< 0.001	
INTERSECTION_DENSITY	7.53E+03	2.12E+03	3.55356	< 0.001	
SEGMENT_80KMHR	-8.64E-04	1.48E-04	-5.82906	< 0.001	
LOW_DENSITY_RESIDENTIAL_AREA_PROP	-5.06E-01	2.07E-01	-2.43935	0.0147	

Model 4

$$TOT = N \times \exp(-1.96) \times$$
$$\exp\left(\begin{array}{l} (\log VKMT \times 0.466) + (INTERSECTION\_DENSITY \times 6.65 \times 10^{3}) + (SEGMENT\_80KMHR \times -8.68 \times 10^{-4}) + \\ (POPULATION\_45TO64 \times -7.16 \times 10^{-4}) + (FOUR\_LEG\_INTERSECTIONS \times 0.0451) \end{array} \right)$$

Model 4					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-1.96E+00	5.32E-01	-3.68906	< 0.001	Dispersion Parameter = 1.3785 Standard Error = 0.135 Log-likelihood = -2516.194
logVKMT	4.66E-01	5.69E-02	8.19349	< 0.001	
INTERSECTION_DENSITY	6.65E+03	2.12E+03	3.14137	< 0.001	
SEGMENT_80KMHR	-8.68E-04	1.49E-04	-5.82509	< 0.001	
POPULATION_45TO64	-7.16E-04	3.36E-04	-2.13409	< 0.001	
FOUR_LEG_INTERSECTIONS	4.56E-02	1.14E-02	3.99918	< 0.001	

Model 5

$$TOT = N \times \exp(-2.25) \times$$
$$\exp\left(\begin{array}{l} (\log VKMT \times 0.501) + (SEGMENT\_80KMHR \times -8.63 \times 10^{-4}) + (FOUR\_LEG\_INTERSECTIONS \times 0.0361) + \\ (INTERSECTION\_DENSITY \times 6.9 \times 10^{3}) + (THREE\_LEG\_INTERSECTIONS \times -0.0169) \end{array} \right)$$

Model 5					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-2.25E+00	5.64E-01	-3.99175	< 0.001	Dispersion Parameter = 1.3871 Standard Error = 0.136 Log-likelihood = -2515.005
logVKMT	5.01E-01	6.11E-02	8.18570	< 0.001	
SEGMENT_80KMHR	-8.63E-04	1.48E-04	-5.81607	< 0.001	
FOUR_LEG_INTERSECTIONS	3.61E-02	1.04E-02	3.48493	< 0.001	
INTERSECTION_DENSITY	6.90E+03	2.14E+03	3.22188	0.0013	
THREE_LEG_INTERSECTIONS	-1.69E-02	6.69E-03	-2.52689	0.0115	

Model 6

$$TOT = N \times \exp(-1.85) \times$$

$$\exp\left(\begin{array}{l}(\log VKMT \times 0.441) + (THREE\_LEG\_INTERSECTIONS \times -0.0105) + (POPULATION\_45TO64 \times 2.01 \times 10^{-5}) + (LOW\_DENSITY\_ \\ RESIDENTIAL\_AREA \times -0.343) + (FOUR\_LEG\_INTERSECTIONS \times 0.0472) + (INTERSECTION\_DENSITY \times 9.79 \times 10^{3})\end{array}\right)$$

Model 6					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-1.85E+00	5.97E-01	-3.10716	0.0019	Dispersion Parameter = 1.2277 Standard Error = 0.118 Log-likelihood = -2541.497
logVKMT	4.41E-01	6.43E-02	6.84666	< 0.001	
THREE_LEG_INTERSECTIONS	-1.05E-02	9.07E-03	-1.15650	0.2475	
POPULATION_45TO64	2.01E-05	5.07E-04	0.03958	0.9684	
LOW_DENSITY_RESIDENTIAL_AREA_PROP	-3.43E-01	2.78E-01	-1.23045	0.2185	
FOUR_LEG_INTERSECTIONS	4.72E-02	1.26E-02	3.74002	< 0.001	
INTERSECTION_DENSITY	9.79E+03	2.24E+03	4.37190	< 0.001	

Model 7

$$TOT = N \times \exp(-3.05) \times$$

$$\exp\left(\begin{array}{l}(\log VKMT \times 0.59) + (I3WP \times -4.36 \times 10^{3}) + (ALKP \times -6.19 \times 10^{3}) + (INTKD \times 0.201) + (FOUR\_LEG\_INTERSECTIONS \times 0.0472) + \\ (SEGMENT\_80KMHR \times -8.11 \times 10^{-4}) + (THREE\_LEG\_INTERSECTIONS \times -0.0214)\end{array}\right)$$

Model 7					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-3.05E+00	5.51E-01	-5.52629	< 0.001	Dispersion Parameter = 1.5197 Standard Error = 0.152 Log-likelihood = -2495.023
logVKMT	5.90E-01	6.33E-02	9.32269	< 0.001	
I3WP	-4.36E-03	2.13E-03	-2.04730	0.0406	
ALKP	-6.19E-03	2.55E-03	-2.43143	0.0150	
INTKD	2.01E-01	3.49E-02	5.76664	< 0.001	
FOUR_LEG_INTERSECTIONS	1.92E-02	1.13E-02	1.69767	0.0896	
SEGMENT_80KMHR	-8.11E-04	1.47E-04	-5.51043	< 0.001	
THREE_LEG_INTERSECTIONS	-2.14E-02	7.66E-03	-2.79157	0.0052	

Model 8

$$TOT = N \times \exp(-1.73) \times$$

$$\exp\begin{pmatrix} (\log VKMT \times 0.458) + (I3WP \times -3.58 \times 10^{-3}) + (INTKD \times 0.121) + (URBAN\_HOLDING\_RESIDENTIAL\_AREA\_PROP \times -2.31) + \\ (FOUR\_LEG\_INTERSECTIONS \times 0.0361) + (SEGMENT\_80KMHR \times -2.94 \times 10^{-4}) + \left(LOW\_DENSITY\_RESIDENTIAL\_AREA \times -0.47\right) \end{pmatrix}$$

Model 8					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-1.73E+00	5.39E-01	-3.20219	0.0014	Dispersion Parameter = 1.827 Standard Error = 0.186 Log-likelihood = -2451.082
logVKMT	4.58E-01	5.93E-02	7.70868	< 0.001	
I3WP	-3.58E-03	1.87E-03	-1.91043	0.0561	
INTKD	1.21E-01	3.32E-02	3.63535	< 0.001	
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	-2.31E+00	2.80E-01	-8.26933	< 0.001	
FOUR_LEG_INTERSECTIONS	3.61E-02	1.11E-02	3.25450	0.0011	
SEGMENT_80KMHR	-2.94E-04	1.44E-04	-2.04095	0.0413	
LOW_DENSITY_RESIDENTIAL_AREA_PROP	-4.74E-01	1.88E-01	-2.52487	0.0116	

Model 9

$$TOT = N \times \exp(-1.7) \times$$

$$\exp\begin{pmatrix} (\log VKMT \times 0.458) + (SEGMENT\_80KMHR \times -2.81 \times 10^{-4}) + (FOUR\_LEG\_INTERSECTIONS \times 0.0368) + \\ \left(INTERSECTION\_DENSITY \times 4.64 \times 10^{3}\right) + (THREE\_LEG\_INTERSECTIONS \times -0.0173) + \\ (ALKP \times 1.17 \times 10^{-3}) + (URBAN\_HOLDING\_RESIDENTIAL\_AREA\_PROP \times -2.54) \end{pmatrix}$$

Model 9					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-1.70E+00	5.56E-01	-3.06063	0.0022	Dispersion Parameter = 1.7684 Standard Error = 0.179 Log-likelihood = -2458.378
logVKMT	4.58E-01	6.10E-02	7.51121	< 0.001	
SEGMENT_80KMHR	-2.81E-04	1.45E-04	-1.93570	0.0529	
FOUR_LEG_INTERSECTIONS	3.68E-02	9.42E-03	3.90467	< 0.001	
INTERSECTION_DENSITY	4.64E+03	2.07E+03	2.24369	0.0249	
THREE_LEG_INTERSECTIONS	-1.73E-02	6.31E-03	-2.74496	0.0061	
ALKP	1.17E-03	2.42E-03	0.48153	0.6301	
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	-2.54E+00	2.75E-01	-9.24670	< 0.001	

Model 10

$$TOT = N \times \exp(-1.74) \times$$
$$\exp\left(\begin{array}{l} (\log VKMT \times 0.466) + (SEGMENT\_80KMHR \times -2.94 \times 10^{-4}) + (FOUR\_LEG\_INTERSECTIONS \times 0.0357) + \\ \left(INTERSECTION\_DENSITY \times 4.68 \times 10^{3}\right) + (THREE\_LEG\_INTERSECTIONS \times -0.0182) + \\ (URBAN\_HOLDING\_RESIDENTIAL\_AREA\_PROP \times -2.5) \end{array} \right)$$

Model 10					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-1.74E+00	5.48E-01	-3.17723	0.0015	Dispersion Parameter = 1.7685 Standard Error = 0.179 Log-likelihood = -2548.582
logVKMT	4.66E-01	5.85E-02	7.96057	< 0.001	
SEGMENT_80KMHR	-2.94E-04	1.45E-04	-2.02392	0.0430	
FOUR_LEG_INTERSECTIONS	3.57E-02	9.20E-03	3.87507	< 0.001	
INTERSECTION_DENSITY	4.68E+03	2.06E+03	2.26850	0.0233	
THREE_LEG_INTERSECTIONS	-1.82E-02	5.95E-03	-3.06244	0.0022	
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	-2.50E+00	2.72E-01	-9.19870	< 0.001	

H2. Fatal-Injury (FI) Collisions Models

Model 1

$$FI = N \times \exp(-3.83) \times \exp\left(\begin{array}{l} (\log VKMT \times 0.527) + (SEGMENT\_80KMHR \times -5.08 \times 10^{-4}) + (FOUR\_LEG\_INTERSECTIONS \times 0.0243) + \\ (\frac{INTERSECTION\_DENSITY}{10000} \times 0.629) + (THREE\_LEG\_INTERSECTIONS \times -0.0221) + \\ (URBAN\_HOLDING\_RESIDENTIAL\_AREA\_PROP \times -2.51) \end{array} \right)$$

Model 1					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-3.83E+00	5.70E-01	-6.71974	< 0.001	Dispersion Parameter = 2.032 Standard Error = 0.234 Log-likelihood = -1744.933
logVKMT	5.27E-01	6.06E-02	8.69232	< 0.001	
SEGMENT_80KMHR	-5.08E-04	1.69E-04	-2.99890	0.0027	
FOUR_LEG_INTERSECTIONS	2.43E-02	8.85E-03	2.74141	0.0061	
INTERSECTION_DENSITY	6.29E-01	2.17E-01	2.90303	0.0037	
THREE_LEG_INTERSECTIONS	-2.21E-02	5.83E-03	-3.78672	< 0.001	
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	-2.51E+00	3.62E-01	-6.93953	< 0.001	

Model 2

$$FI = N \times \exp(-4.3) \times \exp\left(\begin{array}{l} (\log VKMT \times 0.559) + (SEGMENT\_80KMHR \times -9.3 \times 10^{-4}) + (FOUR\_LEG\_INTERSECTIONS \times 0.026) + \\ (\frac{INTERSECTION\_DENSITY}{10000} \times 0.789) + (THREE\_LEG\_INTERSECTIONS \times -0.0211) \end{array} \right)$$

Model 2					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-4.30E+00	5.67E-01	-7.58240	< 0.001	Dispersion Parameter = 1.6699 Standard Error = 0.191 Log-likelihood = -1792.793
logVKMT	5.59E-01	6.10E-02	9.16159	< 0.001	
SEGMENT_80KMHR	-9.30E-04	1.67E-04	-5.55460	< 0.001	
FOUR_LEG_INTERSECTIONS	2.60E-02	9.66E-03	2.69137	0.0071	
INTERSECTION_DENSITY	7.89E-01	2.11E-01	3.73121	< 0.001	
THREE_LEG_INTERSECTIONS	-2.11E-02	6.35E-03	-3.31693	< 0.001	

Model 3

$$FI = N \times \exp(-3.94) \times$$

$$\exp\left(\begin{array}{l} (\log VKMT \times 0.518) + (FOUR\_LEG\_INTERSECTIONS \times 0.0413) + (\dfrac{INTERSECTION\_DENSITY}{10000} \times 0.755) + \\ (SEGMENT\_80KMHR \times -9.7 \times 10^{-4}) + (SEGMENT\_60KMHR \times 3.62 \times 10^{-4}) + (POPULATION\_45TO64 \times -1.11 \times 10^{-3}) \end{array} \right)$$

Model 3					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-3.94E+00	5.36E-01	-7.34862	< 0.001	Dispersion Parameter = 1.6828 Standard Error = 0.192 Log-likelihood = -1790.588
logVKMT	5.18E-01	5.69E-02	9.09223	< 0.001	
FOUR_LEG_INTERSECTIONS	4.13E-02	1.06E-02	3.90220	< 0.001	
INTERSECTION_DENSITY	7.55E-01	2.08E-01	3.63616	< 0.001	
SEGMENT_80KMHR	-9.70E-04	1.72E-04	-5.64123	< 0.001	
SEGMENT_60KMHR	3.62E-04	3.56E-04	1.01606	< 0.001	
POPULATION_45TO64	-1.11E-03	3.15E-04	-3.53152	< 0.001	

Model 4

$$FI = N \times \exp(-4.02) \times$$

$$\exp\left(\begin{array}{l} (\log VKMT \times 0.527) + (\dfrac{INTERSECTION\_DENSITY}{10000} \times 0.768) + (SEGMENT\_80KMHR \times -9.5 \times 10^{-4}) + \\ (POPULATION\_45TO64 \times -1.09 \times 10^{-3}) + (FOUR\_LEG\_INTERSECTIONS \times 0.04) \end{array} \right)$$

Model 4					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-4.02E+00	5.33E-01	-7.54726	< 0.001	Dispersion Parameter = 1.6719 Standard Error = 0.191 Log-likelihood = -1791.674
logVKMT	5.27E-01	5.65E-02	9.32205	< 0.001	
INTERSECTION_DENSITY	7.68E-01	2.08E-01	3.69459	< 0.001	
SEGMENT_80KMHR	-9.50E-04	1.69E-04	-5.62065	< 0.001	
POPULATION_45TO64	-1.09E-03	3.15E-04	-3.45378	< 0.001	
FOUR_LEG_INTERSECTIONS	4.00E-02	1.06E-02	3.78739	< 0.001	

Model 5

$$FI = N \times \exp(-4.37) \times$$

$$\exp\left(\begin{array}{l} (\log VKMT \times 0.57) + (SEGMENT\_80KMHR \times -8.9 \times 10^{-4}) + (FOUR\_LEG\_INTERSECTIONS \times 0.0302) + \\ (\frac{INTERSECTION\_DENSITY}{10000} \times 0.767) + (THREE\_LEG\_INTERSECTIONS \times -0.0131) + \\ \left(\log \frac{COLLECTOR\_LENGTH}{10000} \times -2.39 \right) \end{array} \right)$$

Model 5					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-4.37E+00	5.59E-01	-7.80667	< 0.001	Dispersion Parameter = 1.7232 Standard Error = 0.199 Log-likelihood = -1787.282
logVKMT	5.70E-01	6.02E-02	9.46777	< 0.001	
SEGMENT_80KMHR	-8.91E-04	1.64E-04	-5.44576	< 0.001	
FOUR_LEG_INTERSECTIONS	3.02E-02	9.66E-03	3.12896	0.0018	
INTERSECTION_DENSITY	7.67E-01	2.09E-01	3.67146	< 0.001	
THREE_LEG_INTERSECTIONS	-1.31E-02	7.16E-03	-1.82603	0.0678	
log1p(COLLECTOR_LENGTH)	-2.39E+00	9.89E-01	-2.41718	0.0156	

Model 6

$$FI = N \times \exp(-3.98) \times$$

$$\exp\left(\begin{array}{l} (\log VKMT \times 0.511 + (THREE\_LEG\_INTERSECTIONS \times -5.07 \times 10^{-3}) + (POPULATION\_45TO64 \times 4.52 \times 10^{-5}) + \\ (LOW\_DENSITY\_RESIDENTIAL\_PROP \times -0.417) + (FOUR\_LEG\_INTERSECTIONS \times 0.0431) + \\ \left(\frac{COLLECTOR\_LENGTH}{10000} \times -2.29 \right) + (\frac{INTERSECTION\_DENSITY}{10000} \times 1.03) \end{array} \right)$$

Model 6					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-3.98E+00	5.83E-01	-6.82561	< 0.001	Dispersion Parameter = 1.5071 Standard Error = 0.169 Log-likelihood = -1812.103
logVKMT	5.11E-01	6.23E-02	8.19655	< 0.001	
THREE_LEG_INTERSECTIONS	-5.07E-03	8.58E-03	-0.59125	0.5544	
POPULATION_45TO64	4.52E-05	5.04E-04	0.08968	0.9285	
LOW_DENSITY_RESIDENTIAL_AREA_PROP	-4.17E-01	2.64E-01	-1.57997	0.1141	
FOUR_LEG_INTERSECTIONS	4.31E-02	1.17E-02	3.69264	< 0.001	
COLLECTOR_LENGTH	-2.29E+00	1.04E+00	-2.20421	0.0275	
INTERSECTION_DENSITY	1.03E+00	2.14E-01	4.83324	< 0.001	

Model 7

$$FI = N \times \exp(-3.19) \times \exp\left(\begin{array}{l} (\log VKMT \times 0.455) + (FOUR\_LEG\_INTERSECTIONS \times 0.0396) + (\frac{INTERSECTION\_DENSITY}{10000} \times 0.528) + \\ (SEGMENT\_80KMHR \times -6.2 \times 10^{-4}) + (SEGMENT\_60KMHR \times 1.11 \times 10^{-3}) + \\ (URBAN\_HOLDING\_RESIDENTIAL\_AREA\_PROP \times -2.82) + (POPULATION\_45TO64 \times -1.03 \times 10^{-3}) \end{array} \right)$$

Model 7					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-3.19E+00	5.39E-01	-5.91534	< 0.001	Dispersion Parameter = 2.1123 Standard Error = 0.243 Log-likelihood = -1736.414
logVKMT	4.55E-01	5.65E-02	8.04964	< 0.001	
FOUR_LEG_INTERSECTIONS	3.96E-02	9.53E-03	4.15191	< 0.001	
INTERSECTION_DENSITY	5.28E-01	2.14E-01	2.46761	0.01360	
SEGMENT_80KMHR	-6.20E-04	1.72E-04	-3.59915	< 0.001	
SEGMENT_60KMHR	1.11E-03	3.49E-04	3.19407	< 0.001	
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	-2.82E+00	3.91E-01	-7.21357	< 0.001	
POPULATION_45TO64	-1.03E-03	2.85E-04	-3.59185	< 0.001	

Model 8

$$FI = N \times \exp(-3.75) \times$$
$$\exp\begin{pmatrix} (\log VKMT \times 0.531) + (I3WP \times -5.0 \times 10^{-3}) + (INTKD \times 0.105) + (URBAN\_HOLDING\_RESIDENTIAL\_AREA\_PROP \times -2.4) + \\ (FOUR\_LEG\_INTERSECTIONS \times 0.0251) + (SEGMENT\_80KMHR \times -5.15 \times 10^{-4}) + \\ (LOW\_DENSITY\_RESIDENTIAL\_AREA\_PROP \times -0.549) \end{pmatrix}$$

Model 8					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-3.75E+00	5.72E-01	-6.55296	< 0.001	Dispersion Parameter = 2.0593 Standard Error = 0.238 Log-likelihood = -1742.144
logVKMT	5.31E-01	6.30E-02	8.43085	< 0.001	
I3WP	-5.00E-03	1.98E-03	-2.52449	0.0116	
INTKD	1.05E-01	3.34E-02	3.14198	0.0017	
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	-2.40E+00	3.66E-01	-6.54777	< 0.001	
FOUR_LEG_INTERSECTIONS	2.51E-02	1.11E-02	2.25702	0.0240	
SEGMENT_80KMHR	-5.15E-04	1.70E-04	-3.03384	0.0024	
LOW_DENSITY_RESIDENTIAL_AREA_PROP	-5.49E-01	1.88E-01	-2.92531	0.0034	

Model 9

$$FI = N \times \exp(-3.88) \times$$
$$\exp\begin{pmatrix} (\log VKMT \times 0.537) + (SEGMENT\_80KMHR \times -4.6 \times 10^{-4}) + (FOUR\_LEG\_INTERSECTIONS \times 0.0295) + \\ (\frac{INTERSECTION\_DENSITY}{10000} \times 0.588) + (THREE\_LEG\_INTERSECTIONS \times -0.0127) + \left(\log \frac{COLLECTOR\_LENGTH}{10000} \times -2.76\right) + \\ (URBAN\_HOLDING\_RESIDENTIAL\_AREA\_PROP \times -2.56) \end{pmatrix}$$

Model 9					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-3.88E+00	5.60E-01	-6.93255	< 0.001	Dispersion Parameter = 2.1181 Standard Error = 0.246 Log-likelihood = -1736.162
logVKMT	5.37E-01	5.95E-02	9.02628	< 0.001	
SEGMENT_80KMHR	-4.60E-04	1.66E-04	-2.77313	0.0056	
FOUR_LEG_INTERSECTIONS	2.95E-02	8.80E-03	3.34820	< 0.001	
INTERSECTION_DENSITY	5.88E-01	2.14E-01	2.74995	0.0060	
THREE_LEG_INTERSECTIONS	-1.27E-02	6.53E-03	-1.95181	0.0510	
log1p(COLLECTOR_LENGTH)	-2.76E+00	9.03E-01	-3.05062	0.0023	
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	-2.56E+00	3.59E-01	-7.13789	< 0.001	

Model 10

$$FI = N \times \exp(-3.73) \times$$

$$\exp\left(\begin{array}{l} (\log VKMT \times 0.515) + (SEGMENT\_80KMHR \times -4.04 \times 10^{-4}) + (SEGMENT\_70KMHR \times -2.35 \times 10^{-4}) + \\ (FOUR\_LEG\_INTERSECTIONS \times 0.0304) + (\dfrac{INTERSECTION\_DENSITY}{10000} \times 0.671) + (THREE\_LEG\_INTERSECTIONS \times -0.014) + \\ \left(\log \dfrac{COLLECTOR\_LENGTH}{10000} \times -2.6 \right) + (URBAN\_HOLDING\_RESIDENTIAL\_AREA\_PROP \times -2.72) \end{array} \right)$$

Model 10					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-3.73E+00	5.61E-01	-6.65622	< 0.001	Dispersion Parameter = 2.1548 Standard Error = 0.252 Log-likelihood = -1733.239
logVKMT	5.15E-01	5.99E-02	8.59797	< 0.001	
SEGMENT_80KMHR	-4.04E-04	1.67E-04	-2.41651	0.0157	
SEGMENT_70KMHR	2.35E-04	1.36E-04	1.72792	0.0840	
FOUR_LEG_INTERSECTIONS	3.04E-02	8.74E-03	3.48164	< 0.001	
INTERSECTION_DENSITY	6.71E-01	2.14E-01	3.14138	0.0017	
THREE_LEG_INTERSECTIONS	-1.40E-02	6.53E-03	-2.14826	0.0317	
log1p(COLLECTOR_LENGTH)	-2.60E+00	9.00E-01	-2.88413	0.0039	
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	-2.72E+00	3.74E-01	-7.27061	< 0.001	

H3. Property Damage Only (PDO) Collisions Models

Model 1

$$PDO = N \times \exp(-2.01) \times$$

$$\exp\begin{pmatrix} (\log VKMT \times 0.461) + (I3WP \times -5.01 \times 10^{-3}) + (INTKD \times 0.119) + (FOUR\_LEG\_INTERSECTIONS \times 0.0245) + \\ (SEGMENT\_80KMHR \times -3.26 \times 10^{-4}) + (URBAN\_HOLDING\_RESIDENTIAL\_AREA\_PROP \times -2.18) + \end{pmatrix}$$

Model 1					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-2.01E+00	5.29E-01	-3.79997	< 0.001	Dispersion Parameter = 1.9489 Standard Error = 0.202 Log-likelihood = -2339.605
logVKMT	4.61E-01	5.83E-02	7.90659	< 0.001	
I3WP	-5.01E-03	1.77E-03	-2.83363	0.0046	
INTKD	1.19E-01	3.24E-02	3.65800	< 0.001	
FOUR_LEG_INTERSECTIONS	2.45E-02	9.64E-03	2.54033	0.0111	
SEGMENT_80KMHR	-3.26E-04	1.43E-04	-2.27890	0.0227	
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	-2.13E+00	2.78E-01	-7.68255	< 0.001	

Model 2

$$PDO = N \times \exp(-2.17) \times$$

$$\exp\begin{pmatrix} (\log VKMT \times 0.464) + (FOUR\_LEG\_INTERSECTIONS \times 0.04) + (\frac{INTERSECTION\_DENSITY}{10000} \times 0.736) + \\ (SEGMENT\_80KMHR \times -8.49 \times 10^{-4}) + (THREE\_LEG\_INTERSECTIONS \times -0.0128) + (LOW\_DENSITY\_RESIDENTIAL\_ \\ AREA\_PROP \times -0.41) + (POPULATION\_18TO24 \times 5.1 \times 10^{-3}) + (POPULATION\_45TO64 \times -1.65 \times 10^{-3}) \\ \end{pmatrix}$$

Model 2					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-2.17E+00	5.40E-01	-4.02095	< 0.001	Dispersion Parameter = 1.5589 Standard Error = 0.158 Log-likelihood = -2391.855
logVKMT	4.64E-01	5.84E-02	7.94722	< 0.001	
FOUR_LEG_INTERSECTIONS	4.00E-02	1.15E-02	3.48751	< 0.001	
INTERSECTION_DENSITY	7.36E-01	2.04E-01	3.60454	< 0.001	
SEGMENT_80KMHR	-8.49E-04	1.44E-04	-5.90336	< 0.001	
THREE_LEG_INTERSECTIONS	-1.28E-02	8.12E-03	-1.57316	0.1157	
LOW_DENSITY_RESIDENTIAL_AREA_PROP	-4.10E-01	2.53E-01	-1.62098	0.1050	
POPULATION_18TO24	5.10E-03	2.45E-03	2.08317	0.0372	
POPULATION_45TO64	-1.65E-03	9.17E-04	-1.79644	0.0724	

Model 3

$$PDO = N \times \exp(-1.94) \times$$

$$\exp\left(\begin{array}{l} (\log VKMT \times 0.435) + (FOUR\_LEG\_INTERSECTIONS \times 0.0464) + (\frac{INTERSECTION\_DENSITY}{10000} \times 0.725) + \\ (SEGMENT\_80KMHR \times -8.67 \times 10^{-4}) + (LOW\_DENSITY\_RESIDENTIAL\_AREA\_PROP \times -0.469) \end{array} \right)$$

Model 3					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-1.94E+00	5.03E-01	-3.85140	< 0.001	Dispersion Parameter = 1.5119 Standard Error = 0.152 Log-likelihood = -2397.934
logVKMT	4.35E-01	5.33E-02	8.16762	< 0.001	
FOUR_LEG_INTERSECTIONS	4.64E-02	1.06E-02	4.37957	< 0.001	
INTERSECTION_DENSITY	7.25E-01	2.04E-01	3.54548	< 0.001	
SEGMENT_80KMHR	-8.67E-04	1.46E-04	-5.95225	< 0.001	
LOW_DENSITY_RESIDENTIAL_AREA_PROP	-4.69E-01	1.99E-01	-2.35361	0.0186	

Model 4

$$PDO = N \times \exp(-2.12) \times \exp\left(\begin{aligned} &(\log VKMT \times 0.476) + (I3WP \times -3.77 \times 10^{-3}) + (INTKD \times 0.11) + (FOUR\_LEG\_INTERSECTIONS \times 0.031) + (SEGMENT\_80KMHR \\ &\times -3.26 \times 10^{-4}) + (\frac{COLLECTOR\_LENGTH}{10000} \times 1.75) + (URBAN\_HOLDING\_RESIDENTIAL\_AREA\_PROP \times -2.21) \end{aligned} \right)$$

Model 4					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-2.12E+00	5.25E-01	-4.04026	< 0.001	Dispersion Parameter = 1.998 Standard Error = 0.208 Log-likelihood = -2334.002
logVKMT	4.76E-01	5.80E-02	8.21799	< 0.001	
I3WP	-3.77E-03	1.82E-03	-2.07090	0.0384	
INTKD	1.10E-01	3.21E-02	3.42928	< 0.001	
FOUR_LEG_INTERSECTIONS	3.10E-02	9.91E-03	3.12686	0.0018	
SEGMENT_80KMHR	-3.26E-04	1.41E-04	-2.30500	0.0212	
COLLECTOR_LENGTH	-1.75E+00	7.28E-01	-2.39845	0.0165	
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	-2.21E+00	2.76E-01	-7.99322	< 0.001	

Model 5

$$PDO = N \times \exp(-2.42) \times \exp\left(\begin{aligned} &(\log VKMT \times 0.492) + (SEGMENT\_80KMHR \times -8.68 \times 10^{-4}) + (FOUR\_LEG\_INTERSECTIONS \times 0.0378) + \\ &(\frac{INTERSECTION\_DENSITY}{10000} \times 0.68) + (THREE\_LEG\_INTERSECTIONS \times -0.0162) \end{aligned} \right)$$

Model 5					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-2.42E+00	5.45E-01	-4.44320	< 0.001	Dispersion Parameter = 1.5106 Standard Error = 0.252 Log-likelihood = -2398.194
logVKMT	4.92E-01	5.91E-02	8.33346	< 0.001	
SEGMENT_80KMHR	-8.68E-04	1.46E-04	-5.94616	< 0.001	
FOUR_LEG_INTERSECTIONS	3.78E-02	9.94E-03	3.79847	< 0.001	
INTERSECTION_DENSITY	6.80E-01	2.07E-01	3.28924	0.0010	
THREE_LEG_INTERSECTIONS	-1.62E-02	6.43E-03	-2.51375	0.0119	

Model 6

$$PDO = N \times \exp(-2.03) \times$$

$$\exp\left(\begin{array}{l}(\log VKMT \times 0.432) + (THREE\_LEG\_INTERSECTIONS \times -0.011) + (POPULATION\_45TO64 \times 1.15 \times 10^{-4}) + (LOW\_DENSITY\_ \\ RESIDENTIAL\_AREA\_PROP \times 0.336) + (FOUR\_LEG\_INTERSECTIONS \times 0.0478) + (\dfrac{INTERSECTION\_DENSITY}{10000} \times 0.957)\end{array}\right)$$

Model 6					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-2.03E+00	5.79E-01	-3.50288	< 0.001	Dispersion Parameter = 1.3211 Standard Error = 0.129 Log-likelihood = -2426.387
logVKMT	4.32E-01	6.24E-02	6.93372	< 0.001	
THREE_LEG_INTERSECTIONS	-1.10E-02	8.77E-03	-1.25398	0.2099	
POPULATION_45TO64	1.15E-04	4.90E-04	0.23539	0.8139	
LOW_DENSITY_RESIDENTIAL_AREA_PROP	-3.36E-01	2.69E-01	-1.24640	0.2126	
FOUR_LEG_INTERSECTIONS	4.78E-02	1.22E-02	3.92184	< 0.001	
INTERSECTION_DENSITY	9.57E-01	2.17E-01	4.41362	< 0.001	

Model 7

$$PDO = N \times \exp(-3.02) \times$$

$$\exp\left(\begin{array}{l}(\log VKMT \times 0.542) + (I3WP \times -4.4 \times 10^{-3}) + (INTKD \times 0.196) + (FOUR\_LEG\_INTERSECTIONS \times 0.0266) + \\ (SEGMENT\_80KMHR \times -7.93 \times 10^{-4}) + (THREE\_LEG\_INTERSECTIONS \times -0.0152)\end{array}\right)$$

Model 7					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-3.02E+00	5.29E-01	-5.69883	< 0.001	Dispersion Parameter = 1.6398 Standard Error = 0.167 Log-likelihood = -2380.330
logVKMT	5.42E-01	5.94E-02	9.12497	< 0.001	
I3WP	-4.40E-03	2.06E-03	-2.13336	0.0329	
INTKD	1.96E-01	3.30E-02	5.94110	< 0.001	
FOUR_LEG_INTERSECTIONS	2.66E-02	1.06E-02	2.49813	0.0125	
SEGMENT_80KMHR	-7.93E-04	1.45E-04	-5.47517	< 0.001	
THREE_LEG_INTERSECTIONS	-1.52E-02	7.06E-03	-2.14903	0.0316	

Model 8

$$PDO = N \times \exp(-1.86) \times$$

$$\exp\left(\begin{array}{l}(\log VKMT \times 0.443) + (I3WP \times -3.51 \times 10^{-3}) + (INTKD \times 0.124) + (URBAN\_HOLDING\_RESIDENTIAL\_AREA\_PROP \times -2.18) + \\ (FOUR\_LEG\_INTERSECTIONS \times 0.0378) + (SEGMENT\_80KMHR \times -3.41 \times 10^{-4}) + \\ (LOW\_DENSITY\_RESIDENTIAL\_AREA\_PROP \times -0447)\end{array}\right)$$

Model 8					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-1.86E+00	5.20E-01	-3.57808	< 0.001	Dispersion Parameter = 2.0071 Standard Error = 0.210 Log-likelihood = -2333.605
logVKMT	4.43E-01	5.73E-02	7.73169	< 0.001	
I3WP	-3.51E-03	1.81E-03	-1.93722	0.0527	
INTKD	1.24E-01	3.19E-02	3.88672	< 0.001	
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	-2.18E+00	2.75E-01	-7.94025	< 0.001	
FOUR_LEG_INTERSECTIONS	3.78E-02	1.07E-02	3.54489	< 0.001	
SEGMENT_80KMHR	-3.41E-04	1.42E-04	-2.40784	0.0160	
LOW_DENSITY_RESIDENTIAL_AREA_PROP	-4.47E-01	1.80E-01	-2.48110	0.0131	

Model 9

$$PDO = N \times \exp(-1.82) \times$$

$$\exp\left(\begin{array}{l}(\log VKMT \times 0.444) + (SEGMENT\_80KMHR \times -3.35 \times 10^{-4}) + (FOUR\_LEG\_INTERSECTIONS \times 0.0387) + \\ (\frac{INTERSECTION\_DENSITY}{10000} \times 0.451) + (THREE\_LEG\_INTERSECTIONS \times -0.0164) + (ALKP \times 9.47 \times 10^{-4}) \\ + (URBAN\_HOLDING\_RESIDENTIAL\_AREA\_PROP \times -2.41)\end{array}\right)$$

Model 9					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-1.82E+00	5.38E-01	-3.38051	< 0.001	Dispersion Parameter = 1.9253 Standard Error = 0.199 Log-likelihood = -2342.549
logVKMT	4.44E-01	5.90E-02	7.52907	< 0.001	
SEGMENT_80KMHR	-3.35E-04	1.43E-04	-2.33584	0.0195	
FOUR_LEG_INTERSECTIONS	3.87E-02	9.06E-03	4.26958	< 0.001	
INTERSECTION_DENSITY	4.51E-01	2.01E-01	2.25000	0.0244	
THREE_LEG_INTERSECTIONS	-1.64E-02	6.08E-03	-2.70372	0.0069	
ALKP	9.47E-04	2.35E-03	0.40262	0.6872	
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	-2.41E+00	2.71E-01	-8.90099	< 0.001	

Model 10

$$PDO = N \times \exp(-1.85) \times$$

$$\exp\left(\begin{array}{l}(\log VKMT \times 0.45) + (SEGMENT\_80KMHR \times -3.41 \times 10^{-4}) + (FOUR\_LEG\_INTERSECTIONS \times 0.036) + \\ (\dfrac{INTERSECTION\_DENSITY}{10000} \times 0.453) + (THREE\_LEG\_INTERSECTIONS \times -0.0187) + \\ (URBAN\_HOLDING\_RESIDENTIAL\_AREA\_PROP \times -2.39) + (YOUNG\_DRIVERS \times 1.04 \times 10^{-4})\end{array}\right)$$

Model 10					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
Intercept	-1.85E+00	5.31E-01	-3.48385	< 0.001	Dispersion Parameter = 1.9271 Standard Error = 0.252 Log-likelihood = -2342.571
logVKMT	4.50E-01	5.66E-02	7.94680	< 0.001	
SEGMENT_80KMHR	-3.41E-04	1.43E-04	-2.38148	0.0172	
FOUR_LEG_INTERSECTIONS	3.60E-02	1.02E-02	3.52030	< 0.001	
INTERSECTION_DENSITY	4.53E-01	2.00E-01	2.26020	0.0238	
THREE_LEG_INTERSECTIONS	-1.87E-02	7.07E-03	-2.65026	0.0080	
URBAN_HOLDING_RESIDENTIAL_AREA_PROP	-2.39E+00	2.69E-01	-8.87565	< 0.001	
YOUNG_DRIVERS*	1.04E-04	3.01E-04	0.34463	0.7304	

*YOUNG DRIVERS refers to populations aged 1 to 17 and 18 to 24 years.

H4. Violent Crime Models

Model 1

$$VIOLENT\_CRIMES = N \times \exp(-0.763) \times$$
$$\exp\left(\begin{array}{l} (POPULATION\_DENSITY \times 3.61 \times 10^{-4}) + (POPULATION\_18TO24 \times 9.29 \times 10^{-4}) + (POPULATION\_25TO44 \times 1.5 \times 10^{-3}) + \\ (\frac{RETAIL\_SPACE}{10000} \times 0.338) + (\frac{LAND\_USE\_PER\_TAZ}{10000} \times 2.41 \times 10^{3}) + (\frac{HIGH\_DENSITY\_RESIDENTIAL\_AREA}{10000} \times -0.0514) \end{array} \right)$$

Model 1					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-7.63E-01	2.60E-01	-2.9361	0.0033	Dispersion Parameter = 0.8295 Standard Error = 0.0855 Log-likelihood = -1603.885
POPULATION_DENSITY	3.61E-04	5.51E-05	6.54091	<0.001	
POPULATION_18TO24	9.29E-04	3.60E-03	0.25819	0.7963	
POPULATION_25TO44	1.50E-03	1.38E-03	1.08655	0.2772	
RETAIL_SPACE	3.38E-01	6.05E-02	5.58605	<0.001	
LAND_USE_PER_TAZ	2.41E+03	5.94E+02	4.05633	<0.001	
HIGH_DENSITY_RESIDENTIAL_AREA	-5.14E-02	2.72E-02	-1.8909	0.0586	

Model 2

$$VIOLENT\_CRIMES = N \times \exp(-1.03) \times$$
$$\exp\left(\begin{array}{l} (POPULATION\_18TO24 \times 8.77 \times 10^{-3}) + (POPULATION\_25TO44 \times 8.7 \times 10^{-3}) + (TOT\_POP \times -0.296) + (POPULATION\_ \\ DENSITY \times 2.95 \times 10^{-4}) + (\frac{LAND\_USE\_PER\_TAZ}{10000} \times 3.07 \times 10^{3}) + (\frac{OFFICE\_SPACE}{10000} \times 0.209)(\frac{RETAIL\_SPACE}{10000} \times 0.305) \end{array} \right)$$

Model 2					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-1.03E+00	2.52E-01	-4.0707	<0.001	Dispersion Parameter = 0.9217 Standard Error = 0.0969 Log-likelihood = -1583.339
POPULATION_18TO24	8.77E-03	4.63E-03	1.89506	0.0581	
POPULATION_25TO44	8.70E-03	1.95E-03	4.45303	<0.001	
TOT_POP	-2.96E+01	7.86E+00	-3.7638	<0.001	
POPULATION_DENSITY	2.95E-04	5.45E-05	5.40424	<0.001	
LAND_USE_PER_TAZ	3.07E+03	5.67E+02	5.4088	<0.001	
OFFICE_SPACE	2.09E-01	7.42E-02	2.81998	0.0048	
RETAIL_SPACE	3.05E-01	5.83E-02	5.22521	<0.001	

Model 3

$$VIOLENT\_CRIMES = N \times \exp(-1.03) \times$$

$$\exp\left(\begin{array}{l} (\log \frac{COMMERCIAL\_AREA}{10000} \times 0.6) + (POPULATION\_DENSITY \times 2.95 \times 10^{-4}) + (\frac{LOW\_DENSITY\_RESIDENTIAL\_AREA}{10000} \times 9.29 \times 10^{-3}) + \\ (\frac{RETAIL\_SPACE}{10000} \times 0.114) + (POPULATION\_25TO44 \times 9.14 \times 10^{-3}) + (POPULATION\_45TO64 \times -7.23 \times 10^{-3}) \end{array} \right)$$

Model 3					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-2.45E-01	1.52E-01	-1.6098	0.1074	Dispersion Parameter = 1.0704 Standard Error = 0.116 Log-likelihood = -1555.558
log1p(COMMERCIAL_AREA)	6.00E-01	9.75E-02	6.15772	<0.001	
POPULATION_DENSITY	1.80E-04	5.36E-05	3.35995	<0.001	
LOW_DENSITY_RESIDENTIAL_AREA	9.29E-03	7.71E-03	1.20487	0.2283	
RETAIL_SPACE	1.14E-01	6.42E-02	1.77089	0.0766	
POPULATION_25TO44	9.14E-03	1.23E-03	7.42171	<0.001	
POPULATION_45TO64	-7.23E-03	1.35E-03	-5.354	<0.001	

Model 4

$$VIOLENT\_CRIMES = N \times \exp(-0.543) \times$$

$$\exp\left(\begin{array}{l} (\log \frac{COMMERCIAL\_AREA}{10000} \times 0.789) + (POPULATION\_DENSITY \times 4.12 \times 10^{-4}) + (\frac{INDUSTRY\_SPACE}{10000} \times 0.116) + \\ (\frac{LOW\_DENSITY\_RESIDENTIAL\_AREA}{10000} \times 0.0269) + (\frac{RETAIL\_SPACE}{10000} \times 0.0305) \end{array} \right)$$

Model 4					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-5.43E-01	1.84E-01	-2.946	0.0032	Dispersion Parameter = 0.8223 Standard Error = 0.0838 Log-likelihood = -1603.6700
log1p(COMMERCIAL_AREA)	7.89E-01	9.48E-02	8.31439	<0.001	
POPULATION_DENSITY	4.12E-04	5.38E-05	7.66188	<0.001	
INDUSTRY_SPACE	1.16E-01	5.83E-02	1.99042	0.0465	
LOW_DENSITY_RESIDENTIAL_AREA	2.69E-02	4.32E-03	6.20959	<0.001	
OFFICE_SPACE	3.05E-02	7.74E-02	0.39411	0.6935	

Model 5

$$VIOLENT\_CRIMES = N \times \exp(-0.189) \times$$

$$\exp\left(\begin{array}{l} \left(\log(\frac{COMMERCIAL\_AREA}{10000}) \times 0.574\right) + (POPULATION\_DENSITY \times 1.56 \times 10^{-4}) + (POPULATION\_25TO44 \times 8.46 \times 10^{-4}) + \\ (POPULATION\_45TO64 \times -7.56 \times 10^{-3}) + (POPULATION\_18TO24 \times 4.8 \times 10^{-3}) + (\frac{RETAIL\_SPACE}{10000} \times 0.112) \end{array} \right)$$

Model 5					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-1.89E-01	1.46E-01	-1.296	0.19496	Dispersion Parameter = 1.080 Standard Error = 0.118 Log-likelihood = -1554.416
log1p(COMMERCIAL_AREA)	5.74E-01	9.58E-02	5.98939	<0.001	
POPULATION_DENSITY	1.56E-04	5.07E-05	3.06586	0.0022	
POPULATION_25TO44	8.46E-03	1.40E-03	6.06211	<0.001	
POPULATION_45TO64	-7.56E-03	1.35E-03	-5.5864	<0.001	
POPULATION_18TO24	4.80E-03	3.24E-03	1.4813	0.1385	
RETAIL_SPACE	1.12E-01	6.37E-02	1.75139	0.0799	

Model 6

$$VIOLENT\_CRIMES = N \times \exp(-3.87 \times 10^{-3}) \times$$

$$\exp\left(\begin{array}{l} \left(\log(\frac{COMMERCIAL\_AREA}{10000}) \times 0.644 \right) + (POPULATION\_18TO24 \times 4.14 \times 10^{-3}) + (POPULATION\_25TO44 \times 9.54 \times 10^{-3}) + \\ (POPULATION\_45TO64 \times -7.89 \times 10^{-3}) + (\frac{RETAIL\_SPACE}{10000} \times 0.0838) \end{array} \right)$$

Model 6					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-3.87E-03	1.37E-01	-0.0282	0.9775	Dispersion Parameter = 1.0366 Standard Error = 0.112 Log-likelihood = -1562.308
log1p(COMMERCIAL_AREA)	6.44E-01	9.39E-02	6.86358	<0.001	
POPULATION_18TO24	4.14E-03	3.31E-03	1.25176	0.2107	
POPULATION_25TO44	9.54E-03	1.39E-03	6.88091	<0.001	
POPULATION_45TO64	-7.89E-03	1.38E-03	-5.7269	<0.001	
RETAIL_SPACE	8.38E-02	6.41E-02	1.30708	0.1912	

H5. Assault Crime Models

Model 1

$$ASSAULT\_CRIMES = N \times \exp(-0.625) \times$$

$$\exp\left(\begin{array}{l} (RESIDENTIAL\_PROPORTION \times -1.21) + (RETAIL\_SPACE\_PROP \times 4.8) + \\ (\log LOW\_DENSITY\_RESIDENTIAL\_AREA \times 0.0307) + (LAND\_USE\_PER\_TAZ \times 0.262) + \\ (POPULATION\_25TO44 \times 3.72 \times 10^{-3}) + (COMMERCIAL\_AREA \times 5.55 \times 10^{-6}) \end{array} \right)$$

Model 1					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-6.25E-01	2.87E-01	-2.182918	0.0290	Dispersion Parameter = 0.8008
Residential_Proportion	-1.21E+00	3.38E-01	-3.570788	<0.001	Standard Error = 0.0847
RETAIL_SPACE_PROP	4.80E+00	1.09E+00	4.412598	<0.001	Log-likelihood = -1497.187
log1p(LOW_DENSITY_RESIDENTIAL_AREA)	3.07E-02	2.70E-02	1.136782	0.2556	
LAND_USE_PER_TAZ	2.62E-01	7.00E-02	3.742086	<0.001	
POPULATION_25TO44	3.72E-03	5.28E-04	7.04501	<0.001	
COMMERCIAL_AREA	5.55E-06	1.95E-06	2.845508	0.0044	

Model 2

$$ASSAULT\_CRIMES = N \times \exp(-2.15) \times$$
$$\exp\left(\begin{array}{l} (POPULATION\_25TO44 \times 2.73 \times 10^{-3}) + (INDUSTRY\_SPACE \times 1.42 \times 10^{-5}) + (\log POPULATION\_DENSITY \times 0.174) + \\ (RETAIL\_SPACE\_PROP \times 4.51) + (\log LOW\_DENSITY\_RESIDENTIAL\_AREA \times -0.0188) + \\ (LAND\_USE\_PER\_TAZ \times 0.311) + (COMMERCIAL\_AREA \times 7.4 \times 10^{-6}) \end{array} \right)$$

Model 2					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-2.15E+00	4.62E-01	-4.657117	<0.001	Dispersion Parameter = 0.8031
POPULATION_25TO44	2.73E-03	5.49E-04	4.98581	<0.001	Standard Error = 0.0854
INDUSTRY_SPACE	1.42E-05	7.42E-06	1.910359	0.0561	Log-likelihood = -1498.015
log1p(POPULATION_DENSITY)	1.74E-01	6.07E-02	2.866915	0.0041	
RETAIL_SPACE_PROP	4.51E+00	1.08E+00	4.157656	<0.001	
log1p(LOW_DENSITY_RESIDENTIAL_AREA)	-1.88E-02	2.67E-02	-0.702458	0.4824	
LAND_USE_PER_TAZ	3.11E-01	7.17E-02	4.340826	<0.001	
COMMERCIAL_AREA	7.40E-06	1.93E-06	3.82978	<0.001	

Model 3

$$ASSAULT\_CRIMES = N \times \exp(-0.697) \times$$

$$\exp\begin{pmatrix} (POPULATION\_DENSITY \times 1.91 \times 10^{-4}) + (RETAIL\_SPACE \times 2.49 \times 10^{-5}) + (POPULATION\_25TO44 \times 8.74 \times 10^{-3}) + \\ (POPULATION\_45TO64 \times -6.19 \times 10^{-3}) + (LAND\_USE\_PER\_TAZ \times 0.24) + (RESIDENTIAL\_AREA \times -1.38 \times 10^{-6}) \end{pmatrix}$$

Model 3					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-6.97E-01	2.47E-01	-2.819867	0.0048	Dispersion Parameter = 1.0796 Standard Error = 0.123 Log-likelihood = -1448.016
POPULATION_DENSITY	1.91E-04	5.38E-05	3.555379	<0.001	
RETAIL_SPACE	2.49E-05	5.58E-06	4.458263	<0.001	
POPULATION_25TO44	8.74E-03	1.25E-03	6.982416	<0.001	
POPULATION_45TO64	-6.19E-03	1.27E-03	-4.871482	<0.001	
LAND_USE_PER_TAZ	2.40E-01	5.35E-02	4.486385	<0.001	
Residential_Area	-1.38E-06	3.95E-07	-3.493353	<0.001	

Model 4

$$ASSAULT\_CRIMES = N \times \exp(-1.3) \times$$

$$\exp\begin{pmatrix} (\log COMMERCIAL\_AREA \times 0.171) + (INDUSTRY\_SPACE \times 1.53 \times 10^{-5}) + (OFFICE\_SPACE \times 1.01 \times 10^{-5}) + \\ (POPULATION\_25TO44 \times 4.5 \times 10^{-3}) + (POPULATION\_65\_PLUS \times -2.04 \times 10^{-3}) + (\log POPULATION\_DENSITY \times 0.0657) \end{pmatrix}$$

Model 4					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-1.30E+00	3.64E-01	-3.560387	<0.001	Dispersion Parameter = 0.9705 Standard Error = 0.107 Log-likelihood = -1463.419
log1p(COMMERCIAL_AREA)	1.71E-01	1.81E-02	9.405359	<0.001	
INDUSTRY_SPACE	1.53E-05	6.81E-06	2.250335	0.0244	
OFFICE_SPACE	1.01E-05	7.09E-06	1.418949	0.1559	
POPULATION_25TO44	4.50E-03	5.26E-04	8.559792	<0.001	
POPULATION_65_PLUS	-2.04E-03	8.70E-04	-2.340314	0.0193	
Log_Population_Density	6.57E-02	5.62E-02	1.169566	0.2422	

Model 5

$$ASSAULT\_CRIMES = N \times \exp(-0.78) \times$$
$$\exp\begin{pmatrix} (\log COMMERCIAL\_AREA \times 0.121) + (POPULATION\_DENSITY \times 1.16 \times 10^{-4}) + (POPULATION\_25TO44 \times 8.47 \times 10^{-3}) + \\ (POPULATION\_45TO64 \times -6.68 \times 10^{-3}) + (POPULATION\_18TO24 \times 2.99 \times 10^{-3}) + (RETAIL\_SPACE \times 1.72 \times 10^{-5}) \end{pmatrix}$$

Model 5					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-7.80E-01	1.62E-01	-4.823255	<0.001	Dispersion Parameter = 1.1113 Standard Error = 0.127 Log-likelihood = -1440.754
log1p(COMMERCIAL_AREA)	1.21E-01	1.81E-02	6.67243	<0.001	
POPULATION_DENSITY	1.16E-04	5.19E-05	2.224809	0.0261	
POPULATION_25TO44	8.47E-03	1.39E-03	6.08929	<0.001	
POPULATION_45TO64	-6.68E-03	1.35E-03	-4.937962	<0.001	
POPULATION_18TO24	2.99E-03	3.23E-03	0.925903	0.3545	
RETAIL_SPACE	1.72E-05	5.70E-06	3.019782	<0.001	

Model 6

$$ASSAULT\_CRIMES = N \times \exp(-0.687) \times$$
$$\exp\begin{pmatrix} (\log COMMERCIAL\_AREA \times 0.133) + (POPULATION\_18TO24 \times 2.56 \times 10^{-3}) + (POPULATION\_25TO44 \times 9.17 \times 10^{-3}) + \\ (POPULATION\_45TO64 \times -6.83 \times 10^{-3}) + (RETAIL\_SPACE \times 1.57 \times 10^{-5}) \end{pmatrix}$$

Model 6					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-6.87E-01	1.57E-01	-4.377247	<0.001	Dispersion Parameter = 1.0875 Standard Error = 0.124 Log-likelihood = -1444.756
log1p(COMMERCIAL_AREA)	1.33E-01	1.71E-02	7.760902	<0.001	
POPULATION_18TO24	2.56E-03	3.26E-03	0.785874	0.4319	
POPULATION_25TO44	9.17E-03	1.38E-03	6.663024	<0.001	
POPULATION_45TO64	-6.83E-03	1.37E-03	-5.000085	<0.001	
RETAIL_SPACE	1.57E-05	5.70E-06	2.749171	0.0060	

H6. Robbery Crime Models

Model 1

$$ROBBERY\_CRIMES = N \times \exp(-2.27) \times$$
$$\exp\begin{pmatrix} (RESIDENTIAL\_PROPORTION \times -0.537) + (\log COMMERCIAL\_AREA \times 0.199) + (POPULATION\_18TO24 \times -3.07 \times 10^{-3}) + \\ (POPULATION\_25TO44 \times 4.2 \times 10^{-3}) + (INDUSTRY\_SPACE \times 4.04 \times 10^{-6}) + (OFFICE\_SPACE \times 8.76 \times 10^{-6}) \end{pmatrix}$$

Model 1					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-2.27E+00	3.10E-01	-7.3212	<0.001	Dispersion Parameter = 0.9002
Residential_Proportion	-5.37E-01	3.72E-01	-1.44246	0.1492	Standard Error = 0.124
log1p(COMMERCIAL_AREA)	1.99E-01	2.36E-02	8.418258	<0.001	Log-likelihood = -890.886
POPULATION_18TO24	-3.07E-03	3.70E-03	-0.83037	0.4063	
POPULATION_25TO44	4.20E-03	1.42E-03	2.957995	0.0031	
INDUSTRY_SPACE	4.04E-06	7.05E-06	0.573406	0.5664	
OFFICE_SPACE	8.76E-06	7.71E-06	1.13718	0.2555	

Model 2

$$ROBBERY\_CRIMES = N \times \exp(-2.88) \times$$
$$\exp\begin{pmatrix} (POPULATION\_25TO44 \times 2.18 \times 10^{-3}) + (INDUSTRY\_SPACE \times 8.34 \times 10^{-6}) + (\log POPULATION\_DENSITY \times 0.172) + \\ (RETAIL\_SPACE\_PROP \times 3.3) + (\log LOW\_DENSITY\_RESIDENTIAL\_AREA \times -0.0156) + \\ (LAND\_USE\_PER\_TAZ \times 0.114) + (COMMERCIAL\_AREA \times 1.28 \times 10^{-5}) \end{pmatrix}$$

Model 2					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-2.88E+00	5.83E-01	-4.94324	<0.001	Dispersion Parameter = 0.681 Standard Error = 0.0878 Log-likelihood = -932.5360
POPULATION_25TO44	2.18E-03	6.32E-04	3.448898	<0.001	
INDUSTRY_SPACE	8.34E-06	9.30E-06	0.896875	0.3698	
log1p(POPULATION_DENSITY)	1.72E-01	7.63E-02	2.253736	0.0242	
RETAIL_SPACE_PROP	3.30E+00	1.21E+00	2.730771	0.0063	
log1p(LOW_DENSITY_RESIDENTIAL_AREA)	-1.56E-02	3.15E-02	-0.49495	0.6206	
LAND_USE_PER_TAZ	1.14E-01	8.44E-02	1.348223	0.1776	
COMMERCIAL_AREA	1.28E-05	2.15E-06	5.926497	<0.001	

Model 3

$$ROBBERY\_CRIMES = N \times \exp(-1.4) \times$$
$$\exp\begin{pmatrix} (POPULATION\_DENSITY \times 1.85 \times 10^{-4}) + (RETAIL\_SPACE \times 2.61 \times 10^{-5}) + (POPULATION\_25TO44 \times 9.15 \times 10^{-3}) + \\ (POPULATION\_45TO64 \times -6.96 \times 10^{-3}) + (LAND\_USE\_PER\_TAZ \times 0.0861) + (RESIDENTIAL\_AREA \times -1.99 \times 10^{-6}) \end{pmatrix}$$

Model 3					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-1.40E+00	2.97E-01	-4.71817	<0.001	Dispersion Parameter = 0.9392 Standard Error = 0.135 Log-likelihood = -900.022
POPULATION_DENSITY	1.85E-04	6.33E-05	2.929276	0.0034	
RETAIL_SPACE	2.61E-05	6.34E-06	4.113314	<0.001	
POPULATION_25TO44	9.15E-03	1.53E-03	5.991855	<0.001	
POPULATION_45TO64	-6.96E-03	1.63E-03	-4.26836	<0.001	
LAND_USE_PER_TAZ	8.61E-02	6.49E-02	1.326954	0.1845	
Residential_Area	-1.99E-06	6.90E-07	-2.88078	0.0040	

Model 4

$$ROBBERY\_CRIMES = N \times \exp(-3.03) \times$$
$$\exp\begin{pmatrix} (\log COMMERCIAL\_AREA \times 0.212) + (INDUSTRY\_SPACE \times 1.3 \times 10^{-5}) + (OFFICE\_SPACE \times 1.08 \times 10^{-5}) + \\ (POPULATION\_25TO44 \times 3.42 \times 10^{-3}) + (POPULATION\_65\_PLUS \times -2.9 \times 10^{-3}) + (\log POPULATION\_DENSITY \times 0.0822) \end{pmatrix}$$

Model 4					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-3.03E+00	4.81E-01	-6.30718	<0.001	Dispersion Parameter = 0.9658 Standard Error = 0.136 Log-likelihood = -884.056
log1p(COMMERCIAL_AREA)	2.12E-01	2.26E-02	9.359982	<0.001	
INDUSTRY_SPACE	1.30E-05	8.41E-06	1.543141	0.1228	
OFFICE_SPACE	1.08E-05	7.42E-06	1.45695	0.1451	
POPULATION_25TO44	3.42E-03	5.76E-04	5.936046	<0.001	
POPULATION_65_PLUS	-2.90E-03	9.75E-04	-2.97002	0.002978	
Log_Population_Density	8.22E-02	6.97E-02	1.17914	0.238343	

Model 5

$$ROBBERY\_CRIMES = N \times \exp(-2.1) \times$$
$$\exp\begin{pmatrix} (\log COMMERCIAL\_AREA \times 0.102) + (POPULATION\_DENSITY \times 1.69 \times 10^{-4}) + (POPULATION\_25TO44 \times 7.81 \times 10^{-3}) + \\ (POPULATION\_45TO64 \times -5.77 \times 10^{-3}) + (\log RETAIL\_SPACE \times 0.0838) + (RESIDENTIAL\_PROPORTION \times -0.923) \end{pmatrix}$$

Model 5					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-2.10E+00	2.74E-01	-7.6872	<0.001	Dispersion Parameter = 1.199 Standard Error = 0.179 Log-likelihood = -857.458
log1p(COMMERCIAL_AREA)	1.02E-01	2.45E-02	4.13748	<0.001	
POPULATION_DENSITY	1.69E-04	5.50E-05	3.081584	0.0021	
POPULATION_25TO44	7.81E-03	1.41E-03	5.554947	<0.001	
POPULATION_45TO64	-5.77E-03	1.45E-03	-3.96546	<0.001	
log1p(RETAIL_SPACE)	8.38E-02	2.74E-02	3.05868	0.0022	
Residential_Proportion	-9.23E-01	3.47E-01	-2.65975	0.0078	

Model 6

$$ROBBERY\_CRIMES = N \times \exp(-2.22) \times$$
$$\exp\begin{pmatrix}(\log COMMERCIAL\_AREA \times 0.169) + (POPULATION\_18TO24 \times 3.85 \times 10^{-3}) + (POPULATION\_25TO44 \times 7.58 \times 10^{-3}) + \\ (POPULATION\_45TO64 \times -7.35 \times 10^{-3}) + (RETAIL\_SPACE \times 1.33 \times 10^{-5})\end{pmatrix}$$

Model 6					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-2.22E+00	2.09E-01	-10.6059	<0.001	Dispersion Parameter = 1.093 Standard Error = 0.162 Log-likelihood = -873.527
log1p(COMMERCIAL_AREA)	1.69E-01	2.15E-02	7.863565	<0.001	
POPULATION_18TO24	3.85E-03	3.63E-03	1.063128	0.2877	
POPULATION_25TO44	7.58E-03	1.57E-03	4.818846	<0.001	
POPULATION_45TO64	-7.35E-03	1.60E-03	-4.60895	<0.001	
RETAIL_SPACE	1.33E-05	6.01E-06	2.204654	0.0275	

H7. Break and Enter Crime Models

Model 1

$$BREAK\_AND\_ENTER\_CRIMES = N \times \exp(-0.44) \times$$
$$\exp\begin{pmatrix}(RESIDENTIAL\_PROPORTION \times -0.479) + (\log COMMERCIAL\_AREA \times 0.0485) + (POPULATION\_DENSITY \times 2.01 \times 10^{-5}) + \\ (POPULATION\_25TO44 \times 6.58 \times 10^{-3}) + (POPULATION\_45TO64 \times -3.6 \times 10^{-3}) + (OFFICE\_SPACE \times -4.8 \times 10^{-8})\end{pmatrix}$$

Model 1					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	4.40E-01	1.72E-01	2.5613	0.0104	Dispersion Parameter = 1.5828 Standard Error = 0.177 Log-likelihood = -1477.571
Residential_Proportion	-4.79E-01	2.25E-01	-2.1295	0.0332	
log1p(COMMERCIAL_AREA)	4.85E-02	1.52E-02	3.1889	0.0014	
POPULATION_DENSITY	2.01E-05	4.54E-05	0.4425	0.6581	
POPULATION_25TO44	6.58E-03	1.03E-03	6.4185	<0.001	
POPULATION_45TO64	-3.60E-03	1.03E-03	-3.5082	<0.001	
OFFICE_SPACE	-4.80E-08	5.93E-06	-0.0081	0.9935	

Model 2

$$BREAK\_AND\_ENTER\_CRIMES = N \times \exp(-0.228) \times$$
$$\exp\begin{pmatrix} (LAND\_USE\_PER\_TAZ \times 0.174) + (POPULATION\_25TO44 \times 5.38 \times 10^{-3}) + (POPULATION\_DENSITY \times 9.55 \times 10^{-5}) + \\ (POPULATION\_45TO64 \times -4.85 \times 10^{-3}) + (COMMERCIAL\_AREA \times 3.18 \times 10^{-6}) + \\ (LOW\_DENSITY\_RESIDENTIAL\_AREA \times 1.43 \times 10^{-6}) \end{pmatrix}$$

Model 2					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-2.28E-01	1.89E-01	-1.2081	0.2270	Dispersion Parameter = 1.6623 Standard Error = 0.190 Log-likelihood = -1470.322
LAND_USE_PER_TAZ	1.74E-01	4.32E-02	4.0336	<0.001	
POPULATION_25TO44	5.38E-03	1.01E-03	5.3186	<0.001	
POPULATION_DENSITY	9.55E-05	4.30E-05	2.2224	0.0263	
POPULATION_45TO64	-4.85E-03	1.09E-03	-4.4337	<0.001	
COMMERCIAL_AREA	3.18E-06	1.28E-06	2.4918	0.0127	
LOW_DENSITY_RESIDENTIAL_AREA	1.43E-06	6.30E-07	2.2764	0.0228	

Model 3

$$BREAK\_AND\_ENTER\_CRIMES = N \times \exp(-0.0339) \times$$
$$\exp\begin{pmatrix} (\log COMMERCIAL\_AREA \times 0.0389) + (POPULATION\_DENSITY \times 9.92 \times 10^{-5}) + \\ (LOW\_DENSITY\_RESIDENTIAL\_AREA \times 2.28 \times 10^{-6}) + (RETAIL\_SPACE \times 2.22 \times 10^{-5}) + \\ (POPULATION\_25TO44 \times 5.77 \times 10^{-3}) + (POPULATION\_45TO64 \times -4.98 \times 10^{-3}) \end{pmatrix}$$

Model 3					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-3.39E-02	1.28E-01	-0.2652	0.7909	Dispersion Parameter = 1.7954
log1p(COMMERCIAL_AREA)	3.89E-02	1.42E-02	2.7350	0.0062	Standard Error = 0.207
POPULATION_DENSITY	9.92E-05	4.41E-05	2.2492	0.0245	Log-likelihood = -1455.106
LOW_DENSITY_RESIDENTIAL_AREA	2.28E-06	5.96E-07	3.8206	<0.001	
RETAIL_SPACE	2.22E-05	4.60E-06	4.8241	<0.001	
POPULATION_25TO44	5.77E-03	9.65E-04	5.9864	<0.001	
POPULATION_45TO64	-4.98E-03	1.06E-03	-4.6835	<0.001	

Model 4

$$BREAK\_AND\_ENTER\_CRIMES = N \times \exp(-0.537) \times \exp\begin{pmatrix}(\log COMMERCIAL\_AREA \times 0.0926) + (POPULATION\_DENSITY \times 2.23 \times 10^{-4}) + (INDUSTRY\_SPACE \times 2.83 \times 10^{-5}) + \\ (LOW\_DENSITY\_RESIDENTIAL\_AREA \times 3.16 \times 10^{-6}) + (OFFICE\_SPACE \times 2.56 \times 10^{-6})\end{pmatrix}$$

Model 4					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-5.37E-01	1.45E-01	-3.7060	<0.001	Dispersion Parameter = 1.6532
log1p(COMMERCIAL_AREA)	9.26E-02	1.40E-02	6.6191	<0.001	Standard Error = 0.183
POPULATION_DENSITY	2.23E-04	4.01E-05	5.5678	<0.001	Log-likelihood = -1464.571
INDUSTRY_SPACE	2.83E-05	4.10E-06	6.9049	<0.001	
LOW_DENSITY_RESIDENTIAL_AREA	3.16E-06	3.03E-07	10.4309	<0.001	
OFFICE_SPACE	2.56E-06	5.68E-06	0.4515	0.6516	

Model 5

$$BREAK\_AND\_ENTER\_CRIMES = N \times \exp(-0.112) \times \exp\begin{pmatrix}(\log COMMERCIAL\_AREA \times 0.0394) + (POPULATION\_DENSITY \times 2.68 \times 10^{-5}) + (POPULATION\_25TO44 \times 5.89 \times 10^{-3}) + \\ (POPULATION\_45TO64 \times -4.17 \times 10^{-3}) + (POPULATION\_18TO24 \times 3.11 \times 10^{-3}) + (RETAIL\_SPACE \times 1.72 \times 10^{-5})\end{pmatrix}$$

Model 5					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	1.12E-01	1.27E-01	0.8820	0.3778	Dispersion Parameter = 1.6761 Standard Error = 0.191 Log-likelihood = -1467.641
log1p(COMMERCIAL_AREA)	3.94E-02	1.46E-02	2.6921	0.0071	
POPULATION_DENSITY	2.68E-05	4.32E-05	0.6201	0.5352	
POPULATION_25TO44	5.89E-03	1.13E-03	5.2043	<0.001	
POPULATION_45TO64	-4.17E-03	1.10E-03	-3.7927	<0.001	
POPULATION_18TO24	3.11E-03	2.63E-03	1.1811	0.2376	
RETAIL_SPACE	1.72E-05	4.69E-06	3.6655	<0.001	

Model 6

$$BREAK\_AND\_ENTER\_CRIMES = N \times \exp(-0.132) \times$$
$$\exp\left(\begin{array}{l} (\log COMMERCIAL\_AREA \times 0.0426) + (POPULATION\_18TO24 \times 2.99 \times 10^{-3}) + (POPULATION\_25TO44 \times 6.03 \times 10^{-3}) + \\ (POPULATION\_45TO64 \times -4.19 \times 10^{-3}) + (RETAIL\_SPACE \times 1.67 \times 10^{-5}) \end{array} \right)$$

Model 6					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	1.32E-01	1.23E-01	1.0754	0.2822	Dispersion Parameter = 0.9002 Standard Error = 0.124 Log-likelihood = -890.886
log1p(COMMERCIAL_AREA)	4.26E-02	1.37E-02	3.1161	0.0018	
POPULATION_18TO24	2.99E-03	2.63E-03	1.1365	0.2558	
POPULATION_25TO44	6.03E-03	1.11E-03	5.4401	<0.001	
POPULATION_45TO64	-4.19E-03	1.10E-03	-3.8073	<0.001	
RETAIL_SPACE	1.67E-05	4.64E-06	3.5999	<0.001	

H8. Mischief Crime Models

Model 1

$$MISCHIEF\_CRIMES = N \times \exp(-0.258) \times$$
$$\exp\left(\begin{array}{l} (RESIDENTIAL\_PROPORTION \times -0.367) + (\log COMMERCIAL\_AREA \times 0.094) + (INDUSTRY\_SPACE \times 1.21 \times 10^{-5}) + \\ (OFFICE\_SPACE \times 9.82 \times 10^{-6}) + (POPULATION\_25TO44 \times 4.4 \times 10^{-3}) \end{array} \right)$$

Model 1					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	2.58E-01	1.89E-01	1.3625	0.1730	Dispersion Parameter = 1.6075 Standard Error = 0.185 Log-likelihood = -1662.979
Residential_Proportion	-3.67E-01	2.31E-01	-1.5882	0.1122	
log1p(COMMERCIAL_AREA)	9.40E-02	1.43E-02	6.5757	<0.001	
INDUSTRY_SPACE	1.21E-05	4.34E-06	2.7908	0.0053	
OFFICE_SPACE	9.82E-06	5.56E-06	1.7655	0.0775	
POPULATION_25TO44	4.40E-03	3.14E-04	13.9938	<0.001	

Model 2

$$MISCHIEF\_CRIMES = N \times \exp(-0.343) \times$$
$$\exp\left(\begin{array}{l} (LAND\_USE\_PER\_TAZ \times 0.179) + (POPULATION\_25TO44 \times 4.13 \times 10^{-3}) + (POPULATION\_DENSITY \times 1.55 \times 10^{-4}) + \\ (POPULATION\_45TO64 \times -2.4 \times 10^{-3}) \end{array} \right)$$

Model 2					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	3.43E-01	2.03E-01	1.6905	0.0909	Dispersion Parameter = 1.239 Standard Error = 0.137 Log-likelihood = -1710.194
LAND_USE_PER_TAZ	1.79E-01	4.78E-02	3.7390	<0.001	
POPULATION_25TO44	4.13E-03	1.14E-03	3.6146	<0.001	
POPULATION_DENSITY	1.55E-04	4.53E-05	3.4130	<0.001	
POPULATION_45TO64	-2.40E-03	1.12E-03	-2.1400	0.0324	

Model 3

$$MISCHIEF\_CRIMES = N \times \exp(-0.0262) \times$$

$$\exp\begin{pmatrix} (\log COMMERCIAL\_AREA \times 0.0717) + (POPULATION\_DENSITY \times 1.86 \times 10^{-4}) + \\ (LOW\_DENSITY\_RESIDENTIAL\_AREA \times 2.19 \times 10^{-6}) + (RETAIL\_SPACE \times 2.21 \times 10^{-5}) + (POPULATION\_25TO44 \times 4.12 \times 10^{-3}) + \\ (POPULATION\_45TO64 \times -2.71 \times 10^{-3}) \end{pmatrix}$$

Model 3					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-2.62E-02	1.26E-01	-0.2081	0.8352	Dispersion Parameter = 1.8199 Standard Error = 0.215 Log-likelihood = -1643.669
log1p(COMMERCIAL_AREA)	7.17E-02	1.39E-02	5.1572	<0.001	
POPULATION_DENSITY	1.86E-04	4.29E-05	4.3297	<0.001	
LOW_DENSITY_RESIDENTIAL_AREA	2.19E-06	5.84E-07	3.7548	<0.001	
RETAIL_SPACE	2.21E-05	4.52E-06	4.8913	<0.001	
POPULATION_25TO44	4.12E-03	9.41E-04	4.3822	<0.001	
POPULATION_45TO64	-2.71E-03	1.03E-03	-2.6217	0.0087	

Model 4

$$MISCHIEF\_CRIMES = N \times \exp(-0.253) \times$$

$$\exp\begin{pmatrix} (\log COMMERCIAL\_AREA \times 0.0947) + (INDUSTRY\_SPACE \times 1.85 \times 10^{-5}) + (OFFICE\_SPACE \times 1.05 \times 10^{-5}) + \\ (POPULATION\_25TO44 \times 3.74 \times 10^{-3}) + (POPULATION\_65\_PLUS\_ \times 2.19 \times 10^{-4}) + (\log POPULATION\_DENSITY \times 0.0596) \end{pmatrix}$$

Model 4					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-2.53E-01	2.71E-01	-0.9352	0.3497	Dispersion Parameter = 1.6251 Standard Error = 0.188 Log-likelihood = -1662.476
log1p(COMMERCIAL_AREA)	9.47E-02	1.38E-02	6.8769	<0.001	
INDUSTRY_SPACE	1.85E-05	5.11E-06	3.6181	<0.001	
OFFICE_SPACE	1.05E-05	5.49E-06	1.9057	0.0567	
POPULATION_25TO44	3.74E-03	4.04E-04	9.2464	<0.001	
POPULATION_65_PLUS	2.19E-04	6.67E-04	0.3289	0.7422	
Log_Population_Density	5.96E-02	4.23E-02	1.4083	0.1590	

Model 5

$$MISCHIEF\_CRIMES = N \times \exp(-0.115) \times$$
$$\exp\begin{pmatrix} (\log COMMERCIAL\_AREA \times 0.0658) + (POPULATION\_DENSITY \times 1.21 \times 10^{-4}) + (POPULATION\_25TO44 \times 3.3 \times 10^{-3}) + \\ (POPULATION\_45TO64 \times -2.75 \times 10^{-3}) + (POPULATION\_18TO24 \times 7.52 \times 10^{-3}) + (RETAIL\_SPACE \times 1.98 \times 10^{-5}) \end{pmatrix}$$

Model 5					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	1.15E-01	1.22E-01	0.9419	0.3462	Dispersion Parameter = 1.8149
log1p(COMMERCIAL_AREA)	6.58E-02	1.39E-02	4.7348	<0.001	Standard Error = 0.216
POPULATION_DENSITY	1.21E-04	4.06E-05	2.9871	0.0028	Log-likelihood = -1645.765
POPULATION_25TO44	3.30E-03	1.07E-03	3.0808	0.0021	
POPULATION_45TO64	-2.75E-03	1.04E-03	-2.6449	0.0082	
POPULATION_18TO24	7.52E-03	2.49E-03	3.0221	0.0025	
RETAIL_SPACE	1.98E-05	4.46E-06	4.4507	<0.001	

Model 6

$$MISCHIEF\_CRIMES = N \times \exp(-0.216) \times$$
$$\exp\begin{pmatrix} (\log COMMERCIAL\_AREA \times 0.0792) + (POPULATION\_18TO24 \times 7.01 \times 10^{-3}) + (POPULATION\_25TO44 \times 4.01 \times 10^{-3}) + \\ (POPULATION\_45TO64 \times -2.87 \times 10^{-3}) + (RETAIL\_SPACE \times 1.77 \times 10^{-5}) \end{pmatrix}$$

Model 6					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	2.16E-01	1.20E-01	1.8059	0.0709	Dispersion Parameter = 1.7257
log1p(COMMERCIAL_AREA)	7.92E-02	1.33E-02	5.9677	<0.001	Standard Error = 0.203
POPULATION_18TO24	7.01E-03	2.55E-03	2.7498	0.0060	Log-likelihood = -1653.676
POPULATION_25TO44	4.01E-03	1.07E-03	3.7330	<0.001	
POPULATION_45TO64	-2.87E-03	1.06E-03	-2.6992	<0.001	
RETAIL_SPACE	1.77E-05	4.52E-06	3.9294	<0.001	

H9. Theft Crime Models

Model 1

$$THEFT\_CRIMES = N \times \exp(-0.582) \times$$
$$\exp\begin{pmatrix} (RESIDENTIAL\_PROPORTION \times -1.34) + (\log COMMERCIAL\_AREA \times 0.12) + (POPULATION\_18TO24 \times 5.56 \times 10^{-3}) + \\ (POPULATION\_25TO44 \times 4.53 \times 10^{-3}) + (POPULATION\_45TO64 \times -3.6 \times 10^{-3}) + (RETAIL\_SPACE \times 3.5 \times 10^{-5}) \end{pmatrix}$$

Model 1					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	5.82E-01	1.83E-01	3.1757	0.0015	Dispersion Parameter = 1.3853 Standard Error = 0.153 Log-likelihood = -1610.507
Residential_Proportion	-1.34E+00	2.43E-01	-5.5226	<0.001	
log1p(COMMERCIAL_AREA)	1.20E-01	1.55E-02	7.7532	<0.001	
POPULATION_18TO24	5.56E-03	2.89E-03	1.9208	0.0548	
POPULATION_25TO44	4.53E-03	1.22E-03	3.7088	<0.001	
POPULATION_45TO64	-3.60E-03	1.20E-03	-2.9982	0.0027	
RETAIL_SPACE	3.50E-05	5.16E-06	6.7742	<0.001	

Model 2

$$THEFT\_CRIMES = N \times \exp(-0.66) \times$$
$$\exp\begin{pmatrix} (LAND\_USE\_PER\_TAZ \times 0.0217) + (POPULATION\_25TO44 \times 4.77 \times 10^{-3}) + (POPULATION\_DENSITY \times 1.19 \times 10^{-4}) + \\ (POPULATION\_45TO64 \times -3.71 \times 10^{-3}) + (COMMERCIAL\_AREA \times 1.87 \times 10^{-5}) + \\ (LOW\_DENSITY\_RESIDENTIAL\_AREA \times -2.49 \times 10^{-7}) \end{pmatrix}$$

Model 2					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	6.60E-01	2.39E-01	2.7586	0.0058	Dispersion Parameter = 0.9048 Standard Error = 0.0927 Log-likelihood = -1692.4050
LAND_USE_PER_TAZ	2.17E-02	5.62E-02	0.3866	0.6991	
POPULATION_25TO44	4.77E-03	1.35E-03	3.5339	<0.001	
POPULATION_DENSITY	1.19E-04	5.61E-05	2.1138	0.0345	
POPULATION_45TO64	-3.71E-03	1.46E-03	-2.5461	0.0109	
COMMERCIAL_AREA	1.87E-05	1.66E-06	11.2524	<0.001	
LOW_DENSITY_RESIDENTIAL_AREA	-2.49E-07	8.38E-07	-0.2971	0.7664	

Model 3

$$THEFT\_CRIMES = N \times \exp(-0.111) \times$$

$$\exp\left(\begin{array}{l} (\log COMMERCIAL\_AREA \times 0.134) + (POPULATION\_DENSITY \times 7.99 \times 10^{-5}) + (LOW\_DENSITY\_RESIDENTIAL\_AREA \\ \times -2.36 \times 10^{-7}) + (RETAIL\_SPACE \times 4.61 \times 10^{-5}) + (POPULATION\_25TO44 \times 4.07 \times 10^{-3}) + (POPULATION\_45TO64 \times -2.15 \times 10^{-3}) \end{array} \right)$$

Model 3					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-1.11E-01	1.49E-01	-0.7464	0.4554	Dispersion Parameter = 1.2312 Standard Error = 0.134 Log-likelihood = -1633.262
log1p(COMMERCIAL_AREA)	1.34E-01	1.68E-02	7.9798	<0.001	
POPULATION_DENSITY	7.99E-05	5.18E-05	1.5425	0.1229	
LOW_DENSITY_RESIDENTIAL_AREA	-2.36E-07	7.13E-07	-0.3310	0.7406	
RETAIL_SPACE	4.61E-05	5.42E-06	8.4917	<0.001	
POPULATION_25TO44	4.07E-03	1.15E-03	3.5448	0.0004	
POPULATION_45TO64	-2.15E-03	1.26E-03	-1.6999	0.0892	

Model 4

$$THEFT\_CRIMES = N \times \exp(-0.463) \times$$

$$\exp\left(\begin{array}{l} (\log COMMERCIAL\_AREA \times 0.198) + (INDUSTRY\_SPACE \times 1.94 \times 10^{-5}) + (OFFICE\_SPACE \times 2.35 \times 10^{-5}) + \\ (POPULATION\_25TO44 \times 2.33 \times 10^{-3}) + (POPULATION\_65\_PLUS \times -1.3 \times 10^{-3}) + (\log POPULATION\_DENSITY \times 0.0473) \end{array} \right)$$

Model 4					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-4.63E-01	3.24E-01	-1.4284	0.1532	Dispersion Parameter = 1.0747 Standard Error = 0.113 Log-likelihood = -1657.703
log1p(COMMERCIAL_AREA)	1.98E-01	1.69E-02	11.7228	<0.001	
INDUSTRY_SPACE	1.94E-05	6.18E-06	3.1347	0.0017	
OFFICE_SPACE	2.35E-05	6.65E-06	3.5385	<0.001	
POPULATION_25TO44	2.33E-03	4.98E-04	4.6794	<0.001	
POPULATION_65_PLUS	-1.30E-03	8.25E-04	-1.5818	0.1137	
Log_Population_Density	4.73E-02	5.08E-02	0.9313	0.3517	

Model 5

$$THEFT\_CRIMES = N \times \exp(-0.137) \times$$

$$\exp\left(\begin{array}{l} (\log COMMERCIAL\_AREA \times 0.133) + (POPULATION\_DENSITY \times 9.01 \times 10^{-5}) + (POPULATION\_25TO44 \times 3.1 \times 10^{-3}) + \\ (POPULATION\_45TO64 \times -2.94 \times 10^{-3}) + (POPULATION\_18TO24 \times 4.05 \times 10^{-3}) + (RETAIL\_SPACE \times 4.71 \times 10^{-5}) \end{array} \right)$$

Model 5					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-1.37E-01	1.46E-01	-0.9366	0.3489	Dispersion Parameter = 1.2425 Standard Error = 0.135 Log-likelihood = -1631.520
log1p(COMMERCIAL_AREA)	1.33E-01	1.67E-02	7.9684	<0.001	
POPULATION_DENSITY	9.01E-05	4.87E-05	1.8518	0.0641	
POPULATION_25TO44	3.10E-03	1.30E-03	2.3791	0.0174	
POPULATION_45TO64	-2.94E-03	1.26E-03	-2.3275	0.0199	
POPULATION_18TO24	4.05E-03	3.03E-03	1.3351	0.1818	
RETAIL_SPACE	4.71E-05	5.31E-06	8.8654	<0.001	

Model 6

$$THEFT\_CRIMES = N \times \exp(-0.0634) \times$$

$$\exp\left(\begin{array}{l} (\log COMMERCIAL\_AREA \times 0.142) + (POPULATION\_18TO24 \times 3.67 \times 10^{-3}) + (POPULATION\_25TO44 \times 3.62 \times 10^{-3}) + \\ (POPULATION\_45TO64 \times -3.01 \times 10^{-3}) + (RETAIL\_SPACE \times 4.59 \times 10^{-5}) \end{array} \right)$$

Model 6					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-6.34E-02	1.42E-01	-0.4476	0.6544	Dispersion Parameter = 1.227 Standard Error = 0.133 Log-likelihood = -1634.098
log1p(COMMERCIAL_AREA)	1.42E-01	1.57E-02	9.0452	<0.001	
POPULATION_18TO24	3.67E-03	3.05E-03	1.2047	0.2283	
POPULATION_25TO44	3.62E-03	1.28E-03	2.8180	<0.001	
POPULATION_45TO64	-3.01E-03	1.27E-03	-2.3674	0.0179	
RETAIL_SPACE	4.59E-05	5.29E-06	8.6760	<0.001	

H10. Theft from Auto Crime Models

Model 1

$$THEFT\_FROM\_AUTO\_CRIMES = N \times \exp(-1.34) \times$$
$$\exp\begin{pmatrix}(RESIDENTIAL\_PROPORTION \times -1.06) + (OFFICE\_SPACE \times 1.72 \times 10^{-6}) + (POPULATION\_25TO44 \times 3.13 \times 10^{-3}) + \\ (POPULATION\_65\_PLUS \times 1.8 \times 10^{-3}) + (\log POPULATION\_DENSITY \times -0.0105)\end{pmatrix}$$

Model 1					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	1.34E+00	2.35E-01	5.6876	<0.001	Dispersion Parameter = 1.1723 Standard Error = 0.132 Log-likelihood = -1620.034
Residential_Proportion	-1.06E+00	2.35E-01	-4.5009	<0.001	
OFFICE_SPACE	1.72E-06	6.60E-06	0.2603	0.7946	
POPULATION_25TO44	3.13E-03	4.88E-04	6.4071	<0.001	
POPULATION_65_PLUS	1.80E-03	7.77E-04	2.3164	0.0205	
Log_Population_Density	-1.05E-02	3.68E-02	-0.2857	0.7751	

Model 2

$$THEFT\_FROM\_AUTO\_CRIMES = N \times \exp(-0.726) \times$$
$$\exp\begin{pmatrix}(POPULATION\_25TO44 \times 1.97 \times 10^{-3}) + (INDUSTRY\_SPACE \times 1.96 \times 10^{-5}) + (POPULATION\_DENSITY \times 1.3 \times 10^{-4}) + \\ (RETAIL\_SPACE\_PROP \times 4.19) + (\log LOW\_DENSITY\_RESIDENTIAL\_AREA \times 0.0572) + \\ (LAND\_USE\_PER\_TAZ \times 0.178) + (COMMERCIAL\_AREA \times 1.9 \times 10^{-6})\end{pmatrix}$$

Model 2					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-7.26E-01	2.04E-01	- 3.5624	0.0004	Dispersion Parameter = 1.6087 Standard Error = 0.195 Log-likelihood = -1568.219
POPULATION_25TO44	1.97E-03	3.90E-04	5.0400	<0.001	
INDUSTRY_SPACE	1.96E-05	4.20E-06	4.6684	<0.001	
POPULATION_DENSITY	1.30E-04	4.20E-05	3.0865	<0.001	
RETAIL_SPACE_PROP	4.19E+00	7.72E-01	5.4293	<0.001	
log1p(LOW_DENSITY_RESIDENTIAL_AREA)	5.72E-02	1.88E-02	3.0473	0.0023	
LAND_USE_PER_TAZ	1.78E-01	5.09E-02	3.5007	<0.001	
COMMERCIAL_AREA	1.90E-06	1.38E-06	1.3717	0.1702	

Model 3

$$THEFT\_FROM\_AUTO\_CRIMES = N \times \exp(-0.0415) \times \exp\begin{pmatrix} (\log COMMERCIAL\_AREA \times 0.0601) + (POPULATION\_DENSITY \times 1.34 \times 10^{-4}) + \\ (LOW\_DENSITY\_RESIDENTIAL\_AREA \times 2.85 \times 10^{-6}) + (RETAIL\_SPACE \times 2.62 \times 10^{-5}) + \\ (POPULATION\_25TO44 \times 2.72 \times 10^{-3}) + (POPULATION\_45TO64 \times -1.98 \times 10^{-3})) \end{pmatrix}$$

Model 3					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-4.15E-02	1.34E-01	-0.3108	0.7559	Dispersion Parameter = 1.5886 Standard Error = 0.192 Log-likelihood = -1570.210
log1p(COMMERCIAL_AREA)	6.01E-02	1.49E-02	4.0418	<0.001	
POPULATION_DENSITY	1.34E-04	4.61E-05	2.9018	0.0037	
LOW_DENSITY_RESIDENTIAL_AREA	2.85E-06	6.25E-07	4.5608	<0.001	
RETAIL_SPACE	2.62E-05	4.84E-06	5.4127	<0.001	
POPULATION_25TO44	2.72E-03	1.01E-03	2.6923	0.0071	
POPULATION_45TO64	-1.98E-03	1.11E-03	-1.7914	0.0732	

Model 4

$$THEFT\_FROM\_AUTO\_CRIMES = N \times \exp(-0.115) \times$$
$$\exp\begin{pmatrix}(\log COMMERCIAL\_AREA \times 0.078) + (INDUSTRY\_SPACE \times 2.08 \times 10^{-5}) + (OFFICE\_SPACE \times 4.97 \times 10^{-6}) + \\ (POPULATION\_25TO44 \times 2.96 \times 10^{-3}) + (POPULATION\_65\_PLUS \times 1.14 \times 10^{-3}) + (\log POPULATION\_DENSITY \times 0.0348)\end{pmatrix}$$

Model 4					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-1.15E-01	2.82E-01	-0.4085	0.6829	Dispersion Parameter = 1.4542 Standard Error = 0.173 Log-likelihood = -1584.969
log1p(COMMERCIAL_AREA)	7.80E-02	1.45E-02	5.3608	<0.001	
INDUSTRY_SPACE	2.08E-05	5.34E-06	3.8999	<0.001	
OFFICE_SPACE	4.97E-06	5.86E-06	0.8493	0.3957	
POPULATION_25TO44	2.96E-03	4.28E-04	6.9129	<0.001	
POPULATION_65_PLUS	1.14E-03	7.06E-04	1.6198	0.1053	
Log_Population_Density	3.48E-02	4.43E-02	0.7865	0.4316	

Model 5

$$THEFT\_FROM\_AUTO\_CRIMES = N \times \exp(-0.182) \times$$
$$\exp\begin{pmatrix}(\log COMMERCIAL\_AREA \times 0.0531) + (POPULATION\_DENSITY \times 4.96 \times 10^{-5}) + (POPULATION\_25TO44 \times 2.68 \times 10^{-3}) + \\ (POPULATION\_45TO64 \times -1.13 \times 10^{-3}) + (POPULATION\_18TO24 \times 4.62 \times 10^{-3}) + (RETAIL\_SPACE \times 2.15 \times 10^{-5})\end{pmatrix}$$

Model 5					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	1.82E-01	1.33E-01	1.3635	0.1727	Dispersion Parameter = 1.4642 Standard Error = 0.175 Log-likelihood = -1584.466
log1p(COMMERCIAL_AREA)	5.31E-02	1.54E-02	3.4588	<0.001	
POPULATION_DENSITY	4.96E-05	4.54E-05	1.0925	0.2746	
POPULATION_25TO44	2.68E-03	1.19E-03	2.2449	0.0248	
POPULATION_45TO64	-1.13E-03	1.16E-03	-0.9783	0.3279	
POPULATION_18TO24	4.62E-03	2.78E-03	1.6638	0.0962	
RETAIL_SPACE	2.15E-05	4.95E-06	4.3382	<0.001	

Model 6

$$THEFT\_FROM\_AUTO\_CRIMES = N \times \exp(-0.225) \times$$
$$\exp\begin{pmatrix}(\log COMMERCIAL\_AREA \times 0.0579) + (POPULATION\_18TO24 \times 4.48 \times 10^{-3}) + (POPULATION\_25TO44 \times 2.96 \times 10^{-3}) + \\ (POPULATION\_45TO64 \times -1.2 \times 10^{-3}) + (RETAIL\_SPACE \times 2.07 \times 10^{-5})\end{pmatrix}$$

Model 6					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	2.25E-01	1.29E-01	1.7400	0.0819	Dispersion Parameter = 1.4539 Standard Error = 0.174 Log-likelihood = -1585.626
log1p(COMMERCIAL_AREA)	5.79E-02	1.44E-02	4.0234	<0.001	
POPULATION_18TO24	4.48E-03	2.78E-03	1.6113	0.1071	
POPULATION_25TO44	2.96E-03	1.17E-03	2.5241	0.0116	
POPULATION_45TO64	-1.20E-03	1.16E-03	-1.0352	0.3006	
RETAIL_SPACE	2.07E-05	4.91E-06	4.2111	<0.001	

H11. Theft of Auto Crime Models

Model 1

$$THEFT\_OF\_AUTO\_CRIMES = N \times \exp(-1.19) \times$$
$$\exp\begin{pmatrix}(\log COMMERCIAL\_AREA \times 0.0464) + (POPULATION\_DENSITY \times 2.03 \times 10^{-4}) + (LAND\_USE\_PER\_TAZ \times 0.354) + \\ (URBAN\_HOLDING\_RESIDENTIAL\_AREA \times -7.55 \times 10^{-7})\end{pmatrix}$$

Model 1					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-1.19E+00	2.21E-01	-5.3841	<0.001	Dispersion Parameter = 1.33 Standard Error = 0.158 Log-likelihood = -1388.639
log1p(COMMERCIAL_AREA)	4.64E-02	1.55E-02	2.9923	0.0028	
POPULATION_DENSITY	2.03E-04	4.29E-05	4.7456	<0.001	
LAND_USE_PER_TAZ	3.54E-01	4.21E-02	8.4254	<0.001	
URBAN_HOLDING_RESIDENTIAL_AREA	-7.55E-07	2.81E-07	-2.6911	0.0071	

Model 2

$$THEFT\_OF\_AUTO\_CRIMES = N \times \exp(-1.96) \times$$

$$\exp\begin{pmatrix} (POPULATION\_25TO44 \times 2.65 \times 10^{-3}) + (INDUSTRY\_SPACE \times 2.68 \times 10^{-5}) + (\log POPULATION\_DENSITY \times 0.114) + \\ (RETAIL\_SPACE\_PROP \times 2.14) + (\log LOW\_DENSITY\_RESIDENTIAL\_AREA \times -0.0157) + (LAND\_USE\_PER\_TAZ \times 0.317) + \\ (COMMERCIAL\_AREA \times 3.92 \times 10^{-6}) \end{pmatrix}$$

Model 2					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-1.96E+00	3.61E-01	-5.4232	<0.001	Dispersion Parameter = 1.4793 Standard Error = 0.178 Log-likelihood = -1369.447
POPULATION_25TO44	2.65E-03	4.14E-04	6.3921	<0.001	
INDUSTRY_SPACE	2.68E-05	5.54E-06	4.8387	<0.001	
log1p(POPULATION_DENSITY)	1.14E-01	4.72E-02	2.4153	0.0157	
RETAIL_SPACE_PROP	2.14E+00	8.30E-01	2.5819	<0.001	
log1p(LOW_DENSITY_RESIDENTIAL_AREA)	-1.57E-02	2.06E-02	-0.7605	0.4469	
LAND_USE_PER_TAZ	3.17E-01	5.48E-02	5.7788	<0.001	
COMMERCIAL_AREA	3.92E-06	1.47E-06	2.6626	0.0078	

Model 3

$$THEFT\_OF\_AUTO\_CRIMES = N \times \exp(-1.06) \times$$

$$\exp\begin{pmatrix} (\log COMMERCIAL\_AREA \times 0.0317) + (POPULATION\_DENSITY \times 9.88 \times 10^{-5}) + (RETAIL\_SPACE \times 1.97 \times 10^{-5}) + \\ (POPULATION\_25TO44 \times 5.64 \times 10^{-3}) + (POPULATION\_45TO64 \times -3.88 \times 10^{-3}) + (LAND\_USE\_PER\_TAZ \times 0.238) + \\ (URBAN\_HOLDING\_RESIDENTIAL\_AREA \times -6.09 \times 10^{-7}) \end{pmatrix}$$

Model 3					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-1.06E+00	2.00E-01	-5.3010	<0.001	Dispersion Parameter = 1.9673 Standard Error = 0.261 Log-likelihood = -1332.349
log1p(COMMERCIAL_AREA)	3.17E-02	1.49E-02	2.1231	0.0337	
POPULATION_DENSITY	9.88E-05	4.17E-05	2.3715	0.0177	
RETAIL_SPACE	1.97E-05	4.48E-06	4.4017	<0.001	
POPULATION_25TO44	5.64E-03	9.48E-04	5.9511	<0.001	
POPULATION_45TO64	-3.88E-03	9.46E-04	-4.0996	<0.001	
LAND_USE_PER_TAZ	2.38E-01	4.30E-02	5.5292	<0.001	
URBAN_HOLDING_RESIDENTIAL_AREA	-6.09E-07	2.52E-07	-2.4214	0.0155	

Model 4

$$THEFT\_OF\_AUTO\_CRIMES = N \times \exp(-0.925) \times$$

$$\exp\begin{pmatrix} (\log COMMERCIAL\_AREA \times 0.119) + (INDUSTRY\_SPACE \times 2.55 \times 10^{-5}) + (OFFICE\_SPACE \times -1.02 \times 10^{-6}) + \\ (POPULATION\_25TO44 \times 3.7 \times 10^{-3}) + (POPULATION\_65\_PLUS \times -3.25 \times 10^{-4}) + (\log POPULATION\_DENSITY \times 0.0243) \end{pmatrix}$$

Model 4					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-9.25E-01	2.94E-01	-3.1500	0.0016	Dispersion Parameter = 1.5832 Standard Error = 0.194 Log-likelihood = -1359.952
log1p(COMMERCIAL_AREA)	1.19E-01	1.47E-02	8.1288	<0.001	
INDUSTRY_SPACE	2.55E-05	5.33E-06	4.7811	<0.001	
OFFICE_SPACE	-1.02E-06	5.88E-06	-0.1743	0.8617	
POPULATION_25TO44	3.70E-03	4.18E-04	8.8541	<0.001	
POPULATION_65_PLUS	-3.25E-04	6.88E-04	-0.4716	0.6372	
Log_Population_Density	2.43E-02	4.52E-02	0.5384	0.5903	

Model 5

$$THEFT\_OF\_AUTO\_CRIMES = N \times \exp(-0.403) \times$$

$$\exp\begin{pmatrix} (\log COMMERCIAL\_AREA \times 0.0648) + (POPULATION\_DENSITY \times 5.75 \times 10^{-5}) + (POPULATION\_25TO44 \times 6.07 \times 10^{-3}) + \\ (POPULATION\_45TO64 \times -4.46 \times 10^{-3}) + (POPULATION\_18TO24 \times 3.14 \times 10^{-3}) + (RETAIL\_SPACE \times 1.7 \times 10^{-5}) \end{pmatrix}$$

Model 5					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-4.03E-01	1.36E-01	-2.9645	0.0030	Dispersion Parameter = 1.5991
log1p(COMMERCIAL_AREA)	6.48E-02	1.53E-02	4.2313	<0.001	Standard Error = 0.202
POPULATION_DENSITY	5.75E-05	4.44E-05	1.2941	0.1956	Log-likelihood = -1363.135
POPULATION_25TO44	6.07E-03	1.17E-03	5.1778	<0.001	
POPULATION_45TO64	-4.46E-03	1.14E-03	-3.9196	<0.001	
POPULATION_18TO24	3.14E-03	2.72E-03	1.1546	0.2483	
RETAIL_SPACE	1.70E-05	4.83E-06	3.5152	<0.001	

Model 6

$$THEFT\_OF\_AUTO\_CRIMES = N \times \exp(-0.361) \times$$

$$\exp\left(\begin{array}{l} (\log COMMERCIAL\_AREA \times 0.0721) + (POPULATION\_18TO24 \times 2.78 \times 10^{-3}) + (POPULATION\_25TO44 \times 6.4 \times 10^{-3}) + \\ (POPULATION\_45TO64 \times -4.48 \times 10^{-3}) + (RETAIL\_SPACE \times 1.59 \times 10^{-5}) \end{array} \right)$$

Model 6					
Covariate	**Estimate**	**Standard Error**	**Z value**	**Pr(>|z|)**	**Model Statistics**
(Intercept)	-3.61E-01	1.32E-01	-2.7413	0.0061	Dispersion Parameter = 1.5807
log1p(COMMERCIAL_AREA)	7.21E-02	1.44E-02	4.9966	<0.001	Standard Error = 0.199
POPULATION_18TO24	2.78E-03	2.74E-03	1.0147	0.3103	Log-likelihood = -1364.745
POPULATION_25TO44	6.40E-03	1.15E-03	5.5515	<0.001	
POPULATION_45TO64	-4.48E-03	1.14E-03	-3.9149	<0.001	
RETAIL_SPACE	1.59E-05	4.80E-06	3.3069	<0.001	